

Université de Montréal

Caractérisation génétique d'une forme d'ataxie tardive.

Par
Samira Menasria

Biologie Moléculaire
Faculté de Médecine

Mémoire présenté à la Faculté de Médecine
en vue de l'obtention du grade de maîtrise
en Biologie Moléculaire

avril, 2013

© Samira Menasria, 2013

Résumé

Les ataxies forment un groupe de maladies neurodégénératives qui sont caractérisées par un manque de coordination des mouvements volontaires. Mes travaux ont porté sur une forme d'ataxie à début tardif (LOCA), après l'âge de 50 ans. Les principales caractéristiques cliniques sont: atrophie cérébelleuse à l'IRM (88%), dysarthrie (81%), atrophie du lobe frontal (50%) et nystagmus (52%). La ségrégation dans les familles de cette ataxie est en faveur d'une transmission récessive. Afin d'identifier le gène responsable de LOCA, nous avons recruté 38 patients affectés d'une forme tardive d'ataxie, issus du SLSJ, des Cantons de l'Est ou d'autres régions du Québec. Un premier criblage du génome a été effectué avec des marqueurs microsatellites sur une famille clé. Une analyse de liaison paramétrique nous a suggéré une liaison au chromosome 13 (4.4Mb). Une recherche d'un haplotype partagé entre 17 familles LOCA a diminué la taille de l'intervalle candidat à 1.6Mb, mais l'haplotype s'est avéré fréquent dans la population canadienne-française. Un second criblage du génome avec des marqueurs SNP nous a permis d'évaluer par cartographie d'homozygotie la possibilité qu'une mutation fondatrice partagée dans des sous-groupes de malades. Plusieurs stratégies d'analyse ont été effectuées, entre autre par regroupement régional. Aucun loci candidats ne fut identifié avec confiance. Nous avons donc combiné les données de génotypage avec le séquençage exomique afin d'identifier le gène responsable. L'analyse de six individus atteints nous a permis d'obtenir une liste de variants rare contenant quatre gènes potentiels. Cette analyse doit se poursuivre pour identifier le gène responsable de LOCA.

Mots-clés : Ataxie, neurodégénérative, séquençage exomique, cartographie par homozygotie

Abstract

Ataxias are a heterogeneous group of neurodegenerative diseases and are characterized by a lack of voluntary movements. My Master's project was on a late-onset ataxia (LOCA), after 50 years of age. The main clinical features are: cerebellar atrophy on MRI (88%), dysarthria (81%), frontal lobe atrophy (50%) and nystagmus (52%). Disease segregation in the family is suggestive of a recessive transmission. In order to identify the causal gene of LOCA, we have recruited 38 patients affected by a late-onset ataxia, originated from SLSJ, Eastern townships or other region in Quebec. A first genome scan was done with microsatellite markers on an informative family. Parametric linkage analysis suggested linkage on chromosome 13 (4.4Mb). Haplotype sharing analysis on 17 families reduced the candidate interval to 1.6Mb, but this haplotype was found to be frequent in the French-Canadian population. A second genome scan with SNPs markers allowed us to performed homozygosity mapping and look for founder mutations in subgroup of patients. Many strategies were performed, including regional clustering. No candidate loci were identified with confidence. We decided to combine the genotyping analysis results with exome sequencing to uncover the causative gene. The analysis on six affected individuals allowed us to obtain a rare variants list with four putative genes. More analysis is needed to identify the gene responsible for LOCA.

Keywords: Ataxia, neurodegenerative, exome sequencing, homozygosity mapping

Table des matières

Résumé.....	i
Abstract.....	ii
Table des matières.....	iii
Liste des tableaux.....	v
Liste des figures.....	vi
Liste des abréviations.....	vii
Dédicace.....	ix
Remerciements.....	x
Introduction.....	1
1.1 Les ataxies cérébelleuses.....	1
1.1.1 Ataxies cérébelleuses à transmission autosomique récessive.....	1
1.1.2 Ataxies cérébelleuses à transmission autosomique dominante.....	4
1.1.3 Ataxies cérébelleusesà transmission liée au chromosome X.....	6
1.1.4 Ataxiescérébelleuseà transmission mitochondriale.....	7
1.2 Approche pour identifier des loci candidats et des gènes causant des maladies héréditaires.....	7
1.2.1 Marqueurs génétiques.....	8
1.2.2 Analyses génétique.....	9
1.2.3 Séquençage.....	10
1.3 Projet de caractérisation génétique d'une forme d'ataxie tardive.....	11
Chapitre 2: Identification de loci LOCA candidats.....	12
2.1 Introduction.....	12
2.2 Resultats.....	12
2.2.1 Cohorte LOCA.....	12
2.2.2 Exclusion du locus situé sur le chromosome 13.....	14
2.2.3 Analyse de partage d'homozygotie entre les familles LOCA.....	17
2.2.4 Partage d'homozygotie entre les différentes régions du Québec.....	23

2.2.5 Partage d'haplotype entre deux familles originaires des Cantons de l'Est.	25
2.2.6 Exclusion du gène candidat SYT14.	26
2.3 Conclusion.	27
2.4 Matériels et méthodes.	28
2.4.1 Recrutement des patients.	28
2.4.2 Génotypage et analyse génétique avec marqueurs STS.	28
2.4.3 Génotypage et analyse génétique avec marqueurs SNP.	28
2.4.4 PCR et séquençage du gène SYT14.	29
2.5 Contributions.	30
 Chapitre 3: Recherche de mutations par séquençage exomique.	 31
3.1 Introduction.	31
3.2 Résultats.	31
3.2.1 Séquençage exomique des patients LOCA.	31
3.2.2 Partage de variants entre les six atteints LOCA.	34
3.2.3 Partage de variants dans les régions candidates identifiées par criblage entier du génome	34
3.2.4 Partage de variants dans les régions candidates identifiées par partage d'homozygotie .	36
3.2.5 Partage de variants dans les régions candidates identifiées par partage d'haplotype entre les deux familles des Cantons de l'Est	38
3.2.6 Partage de variants entre les quatre atteints LOCA originaire du SLSJ	39
3.3 Conclusion	41
3.4 Matériels et méthodes	41
3.4.1 Séquençage exomique.	41
3.4.2 Analyse des données d'exome.	41
3.5 Contributions	42
 Chapitre 4: Conclusions et perspectives.	 43
4.1 Conclusions du projet	43
4.2 Perspectives.	45

Liste des tableaux

Table I	Ataxies cérébelleuses récessives
Table II	Ataxies cérébelleuses dominantes
Table III	Résumé de partage d'homozygotie entre les atteints LOCA
Table IV	Liste des priorités d'intérêt des régions d'homozygotie partagées entre les atteints LOCA
Table V	Partage d'haplotype entre deux familles originaires des Cantons de l'Est
Table VI	Variants identifiés dans SYT14 dans la cohorte CF
Table VII	Amorces utilisés pour l'amplification des exons du gène SYT14.

Liste des figures

- Figure 1 Pedigree des familles LOCA
- Figure 2 Distribution géographique des cas LOCA
- Figure 3 Graphique démontrant les résultats de criblage du génome par marqueurs STS
- Figure 4 LOD score multipoint pour les familles LOCA dans la région candidate du chromosome 13
- Figure 5 Analyse de partage d'homozygotie entre les atteints LOCA
- Figure 6 Partage d'homozygotie entre les atteints LOCA
- Figure 7 Nombre de régions homozygotes de de plus de 1Mb de longueur observées selon la region geographique
- Figure 8 Longueur des régions homozygotes (minimum 1Mb) et longueur moyenne des régions homozygotes par atteint LOCA (barre de couleur verte)
- Figure 9 Distribution géographique des atteints LOCA analysés par séquençage exomique
- Figure 10 Pedigree des familles originaires du SLSJ envoyé pour séquençage exomique
- Figure 11 Représentation schématique de l'analyse partage de variants entre les six atteints LOCA.
- Figure 12 Répartition des variants après séquençage exomique dans les loci candidats identifiés par analyse de liaisons de la famille 14.
- Figure 13 Répartition des variants après séquençage exomique dans les loci candidats identifiés par partage d'homozygotie entre les atteints LOCA.
- Figure 14 Répartition des variants après séquençage exomique dans les loci candidats identifié par partage d'haplotype dans deux familles originaire des Cantons de l'Est.
- Figure 15 Répartition des variants après séquençage exomique entre les atteints LOCA selon un mode de transmission autosomique dominant.

Liste des abréviations

ADN Acide désoxyribonucléique

AOA Ataxie oculomotrice avec apraxie

ASRACS Ataxie spastique récessive autosomique de Charlevoix-Saguenay

ATM ataxia telangiectasia mutated

CHU Centre hospitalier universitaire

CHUM Centre hospitalier de l'Université de Montréal

cM Centi Morgan

CRCHUM Centre de recherche du centre hospitalier de l'Université de Montréal

DMOP Dystrophie musculaire oculo-pharyngée

FRDA Friedreich ataxia

FXTAS Fragile X-associated tremor/ataxia syndrome

GWS Genome-wide scan

Kb Kilobase

KCNH1 Potassium channel H1

LOCA late-onset cerebellar ataxia

LOD Logarithm of odds

Mb Mégabase

MSA multisystem atrophy

ng Nanogramme

POLG Polymerase (DNA Directed), Gamma

RT-PCR Reverse transcript PCR

SERTAD4 SEI-1, RBT1 and TARA domain 4

SCA spinocerebellar ataxia

S-L-S-J Saguenay-Lac-Saint-Jean

SNP Single nucleotide polymorphism

STR Short tandem repeat

SYT14 Synaptotagmin 14

U Unité

U/ μ l Unité par microlitre

μl Microlitre

μM Micromolaire

À mes parents que j'aime tant...

Remerciements

Je voudrais tout d'abord remercier mon directeur de recherche, Dr Bernard Brais, de m'avoir si bien accueillie dans son laboratoire et de m'avoir donné l'opportunité de faire de la recherche en me guidant tout au long de mes études qui s'est montré toujours à l'écoute et compréhensif malgré son horaire chargé, qui m'a poussé plus loin et m'a tant appris.

Je voudrais aussi remercier les membres du laboratoire : Martine, Talita, Catherine, Roxanne, Rebecca Kim et Marie-Josée et anciens membres : Véronique, Emmanuelle, Julie qui ont bien voulu partager leurs connaissances avec moi, des personnes qui me sont devenues chères et auprès de qui travailler était un plaisir. Tout particulièrement je voudrais remercier Martine Tétreault qui a grandement contribué à ma formation, qui m'a donné de son temps sans compter, une collègue qui est devenue une amie et m'a épaulé avec patience et extrême gentillesse tout au long de mon projet, un merci serait trop faible!

Je remercie également Richard Bertrand et Vivianne Jodoin pour leur disponibilité à chaque étape de ce cycle supérieur et pour leurs précieux conseils, et aussi les membres du jury qui ont accepté de juger ce travail.

Je voudrais également remercier les patients et leurs familles de nous avoir fait confiance.

J'aimerais enfin remercier mon amoureux, ma chère famille et mes amis pour leur support inconditionnel.

Introduction

- **1.1 Les ataxies cérébelleuses.**

L'ataxie cérébelleuse est un groupe de maladie neurodégénératives avec une hétérogénéité clinique et génétique (Hersheson, Haworth et al. 2012). Ce groupe de maladies est caractérisé par une dégénérescence progressive du cervelet, ce qui résulte en des problèmes d'équilibre, de coordination des mouvements et une dysarthrie. Il y a des formes dites d'ataxie cérébelleuse pure, où les symptômes sont isolés au cervelet, mais les formes les plus fréquentes sont accompagnées d'autres manifestations neurologiques (pyramidales, extrapyramidales, sensitives ou cognitive par exemple) (Hersheson, Haworth et al. 2012). Les ataxies cérébelleuses peuvent être génétique, acquise ou non-génétique (van Gaalen and van de Warrenburg 2012). Les ataxies cérébelleuses héréditaires peuvent être transmises sous plusieurs modes de transmission (autosomique récessive et dominante, liée au chromosome X ainsi que mitochondriale). Les ataxies cérébelleuses sont principalement classées selon le mode de transmission, le gène causal ou le locus chromosomique (Jayadev and Bird 2013). Une deuxième classification est ensuite effectuée selon l'âge de début des symptômes (pédiatrique, adolescence-jeune adulte ou tardif). Le diagnostic des ataxies cérébelleuses est très difficile. Ceci est dû à une hétérogénéité clinique (chevauchement des signes cliniques entre les différents types d'ataxies) et d'une hétérogénéité génétique (plusieurs gènes causent un phénotype similaire) (van Gaalen and van de Warrenburg 2012). De plus, même les formes communes d'ataxies sont considérées comme des maladies rares il est donc difficile d'identifier plusieurs familles avec la même maladie.

- **1.1.1 Ataxies cérébelleuses à transmission autosomique récessive**

Les ataxies cérébelleuses transmises sous le mode autosomique récessif sont très hétérogènes dans leur âge de début des symptômes, la sévérité de la progression et la fréquence des signes non-cérébelleux (Table I) (Anheim, Tranchant et al. 2012). Plus fréquemment, elles sont

associées à un âge de début des symptômes se situant dans l'enfance ou l'adolescence (moins de 20 ans). Une variante moins sévère de ces maladies peut se présenter avec un âge de début plus tardif (Fogel 2012). Les ataxies cérébelleuses récessives sont caractérisées par une progression lente d'une ataxie des membres symétrique avec détérioration progressive de la marche. Elles sont souvent associées à une polyneuropathie sensitive ou sensitivomotrice (Fogel 2012). Jusqu'à maintenant plus d'une centaine de gènes ont été identifiés comme étant responsables de ce groupe de maladies (Hersheson, Haworth et al. 2012).

L'ataxie de Friedreich (FRDA) est l'ataxie récessive la mieux connue et la plus prévalente internationalement (Fogel 2012, van Gaalen and van de Warrenburg 2012). Le phénotype classique se présente avant l'âge de 20 ans (Table I). Le phénotype consiste en une ataxie cérébelleuse progressive affectant la marche, associé à une ataxie sensitive, aréflexie et une faiblesse pyramidale des membres inférieurs (Fogel 2012). Environ 25% des individus atteints de FRDA ont un phénotype atypique, comme par exemple un début tardif de la maladie. Des mutations dans le gène frataxin (*FXN*) ont été décrites pour la première fois en 1996, il s'agissait de l'expansion d'une répétition GAA (Campuzano, Montermini et al. 1996). En fait, 98% des patients atteints de FRDA sont porteur de cette répétition et seulement 2% possède des mutations ponctuelles (Delatycki, Williamson et al. 2000). Aucun patient porteur de deux mutations ponctuelles n'a été décrit à ce jour, mais certains individus sont hétérozygotes composés pour l'expansion et une mutation ponctuelle (Hersheson, Haworth et al. 2012). La mutation commune cause la diminution de l'expression du gène et perturbe le métabolisme du fer qu'effectue la mitochondrie (Fogel 2012).

Parmi les autres ataxies récessives, on retrouve entre autre l'ataxie télangiectasie (ATM). Il s'agit de la deuxième forme la plus fréquente dans le monde, et la première pour les formes avec âge de début avant 5 ans (Fogel 2012). Certains cas atypiques avec un début de la maladie après 30 ans ont été observés (van Gaalen and van de Warrenburg 2012). En plus de l'ataxie à la marche et des membres, les patients sont atteints d'une apraxie oculomotrice et une télangiectasie oculo-cutanée. Elle est causée par des mutations dans le gène *ATM* codant pour une serine/thréonine kinase impliqué dans la réparation de l'ADN. L'ataxie récessive avec apraxie oculomotrice de type 2 (AOA2) est la forme la plus fréquente avec un début des

symptômes dans l'adolescence. Cette forme d'ataxie ressemble à bien des niveaux cliniquement à la FRDA et à ATM (Fogel 2012). Elle se présente avec une ataxie à la marche associée à une neuropathie sensitivomotrice, une apraxie oculomotrice et un niveau d'apha-fetoprotéine élevé. AOA2 est causée par des mutations dans le gène *senataxin* (*SETX*). Bien que des cas atteints de AOA2 aient été décrit partout dans le monde, une prévalence plus élevée de la maladie est observée chez les Canadiens-Français (CF), plus particulièrement de la région de la Gaspésie (Duquette, Roddier et al. 2005). Une mutation fondatrice (p.Leu1986Arg) a été identifiée chez les dix familles étudiées.

D'autres ataxies cérébelleuses récessives plus fréquentes chez les CF ont été décrites (Table I). L'ataxie spinocérébelleuse de type 8 (SCAR8 ou ARCA1) a été décrite pour la première fois dans une cohorte canadienne-française, plus particulièrement originaire de la région de la Beauce et du Bas-St-Laurent (Gros-Louis, Dupre et al. 2007) (Table I). ARCA1 est causée par des mutations dans le gène *SYNE1*. ARCA1 est caractérisé par un début des symptômes à l'âge adulte (entre 17 et 46 ans, moyenne 31 ans), une progression lente de l'ataxie et de la dysarthrie. Aucun signe extrapyramidal, atrophie optique, perte auditive ou anomalie sensitive a été observé chez les patients CF. L'ataxie de Charlevoix-Saguenay (ARSACS) a aussi été décrite pour la première fois dans une cohorte CF, plus particulièrement de la région du Saguenay-Lac-St-Jean (SLSJ) et Charlevoix avec un taux de porteur de 1:22 (Engert, Dore et al. 1999) (Table I). Cette ataxie spastique est caractérisée par un début dans l'enfance de l'ataxie cérébelleuse, des signes pyramidaux et une neuropathie périphérique. Des mutations dans le grand gène *sacsin* (*SACS*), dont une mutation fondatrice (6594delT) chez les CF, ont été associées avec ARSACS. Des patients atteints d'ARSACS ont maintenant été décrit dans plusieurs autres pays, avec plus de 98 mutations dans le gène *SACS* (Thiffault, Dicaire et al. 2013).

Table I: Ataxies cérébelleuses récessives

Disease	OMIM #	Gene	Protein	Mutation	Incidence/carrier frequency	Geographic distribution	Characteristic features
Friedrich ataxia	606829	<i>FXN</i>	Frataxin	GAA repeat expansions; point mutations in compound heterozygotes	Incidence: 1:30–50,000 Carrier frequency: 0.9–1.6%	Worldwide except natives to: Far East, sub-Saharan Africa, Australia, America	Spasticity, neuropathy, cardiac involvement
Ataxia-telangiectasia	607585	<i>ATM</i>	Ataxia telangiectasia mutated	Deletions: splice-site related; nonsense; missense	Incidence: 1:400,000–450,000 live births Carrier frequency: 0.35–1%	Reported in many worldwide populations	Oculomotor apraxia; extrapyramidal features; increased cancer risk/radiosensitivity
Ataxia-telangiectasia like disorder (ATLD)	604391	<i>MRE11A</i>	Meiotic recombination 11, <i>S. cerevisiae</i> , homolog of	Missense	25 reported cases worldwide	Saudi Arabia (15 cases), Japan (4 cases), UK (4 cases), Italy (2 cases)	Similar to ATM but milder phenotype
Ataxia-oculomotor apraxia type 1	208920	<i>APTX</i>	Aprataxin	Insertion, deletion, missense	Rare worldwide—More common in Portuguese and Japanese populations	Portugal, Japan, France, Tunisia	Oculomotor apraxia, peripheral neuropathy
Cerebellar ataxia with muscle coenzyme Q10 deficiency	607426	<i>APTX</i>	Aprataxin	Missense	Rare	Single Italian family	Low coenzyme Q10 levels; late-onset hypergonadotrophic hypogonadism
Ataxia-oculomotor apraxia type 2	606002	<i>SETX</i>	Senataxin	Nonsense, missense	Carrier frequency: 2.1–3.5% Incidence: 1:400,000 (Alsace)	Commoner in French-Canadian populations	Oculomotor apraxia (variable); extrapyramidal features; peripheral neuropathy
Spastic ataxia of Charlevoix-Saguenay (ARSACS)	270550	<i>SACS</i>	Sacsin	Stop-gain deletions and point mutations most common	Carrier frequency (Quebec): 4.5% Incidence: 1/1930	Most common in Quebec, Tunisian, Turkish, Italian, Japanese families reported	Myelinated retinal fibers; prominent lower limb spasticity
Cerebellar ataxia, seizures and ubiquinone deficiency	612016	<i>ADCK3</i>	aarF domain containing kinase 3	Missense, splice site, frame shift, deletion	Rare	French, Dutch, British families reported	Mental retardation, seizures, low coenzyme q10 levels
Spinocerebellar ataxia with axonal neuropathy (SCAN1)	607250	<i>TDP1</i>	Tyrosyl DNA phosphodiesterase 1	Missense	Rare	Saudi Arabian family	Axonal neuropathy
Autosomal recessive spinocerebellar ataxia type 8	610743	<i>SYNE1</i>	Synaptic nuclear envelope protein 1	Splice site, intronic	Rare worldwide 3rd most common ARCA in Quebec	Canada	Hyperreflexia
Autosomal recessive spinocerebellar ataxia type 10	613728	<i>ANO10</i>	Anoctamin 10	Missense, splice site, deletion	Rare	French, Dutch, Serbian families	Tortuous conjunctival vessels

Hersherson J et al Hum Mut 2012 Sep; 33 (9): 1324-32

■ 1.1.2 Ataxies cérébelleuses à transmission autosomique dominante

Les ataxies cérébelleuses transmises de façon dominante sont appelées ataxies spinocérébelleuses (SCA) (Table II) (van Gaalen and van de Warrenburg 2012). Elles sont caractérisées par une atteinte du cervelet, du tronc cérébrale et des voies spinocérébelleuses (Doi, Yoshida et al. 2011). Elles peuvent parfois progresser en un syndrome avec atrophie multi-système (AMS, multisystem atrophy (MSA) en Anglais). L'AMS associé à des signes parkinsoniens est la forme la plus fréquente (van Gaalen and van de Warrenburg 2012). Il y a plus de 30 loci SCA décrits jusqu'à maintenant. Pour plus d'une vingtaine d'entre elles le gène a été identifié (Hersherson, Haworth et al. 2012). Dans la majorité des ataxies cérébelleuses dominantes, l'âge de début des symptômes est tardif et se situe entre 20 et 50 ans (Fogel 2012). Quelques formes plus rares d'ataxies spinocérébelleuses peuvent aussi avoir un début de la maladie dans l'enfance (par exemple: SCA2, SCA7, SCA18, SCA21, SCSA28 et SCA29).

La majorité des ataxies cérébelleuses dominantes sont causées par une expansion d'une séquence répétée codante ou non. Les expansions polyglutamine (CAG) sont les plus

communes et sont retrouvées dans: SCA1, SCA2, SCA3, SCA6, SCA7, SCA17 et DRPLA (Hersheson, Haworth et al. 2012). Une expansion dans une région non-codante est observée pour les formes suivantes: SCA8, SCA10, SCA12, SCA31 et SCA36. Une corrélation phénotype-génotype est bien décrite dans plusieurs de ces formes. Une corrélation entre la grandeur de l'expansion et la sévérité de la maladie est généralement observée. Plus l'expansion est grande plus l'âge de début des symptômes est précoce et la sévérité grande. Malgré tout, la minorité des sous-types des ataxies cérébelleuses dominantes sont causées par des mutations non-répétées conventionnelles, par exemple: SCA5, SCA11, SCA13, SCA14, SCA15, SCA20, SCA23, SCA27, SCA28 et SCA35. Mais dans les faits, 45% des ataxies dominantes sont causées par des expansions de régions répétées, 6% par des mutations conventionnelles et une grande proportion, 48%, sont toujours sans diagnostic (Hersheson, Haworth et al. 2012). SCA1 est en fait le premier gène responsable d'une ataxie à être identifié, l'expansion de la répétition CAG dans le gène *ATXN1* a été observée pour la première fois en 1993 (Orr, Chung et al. 1993).

D'autres ataxies à transmission dominante ont été décrites. L'atrophie dentatorubro pallidoluisienne (DRPLA) est l'une d'entre elles (Fogel 2012, van Gaalen and van de Warrenburg 2012). Il y a aussi les ataxies épisodiques, qui consistent en un groupe de maladies hétérogènes dont huit formes ont été décrites jusqu'à maintenant (EA1 à 7 et CSE). EA1 et EA2 sont les formes les plus communes (Hersheson, Haworth et al. 2012). Les ataxies épisodiques, dont les gènes responsables sont identifiés, sont causées par des désordres au niveau de canaux ioniques. Les patients présentent des attaques d'ataxie qui peuvent être associées à plusieurs autres symptômes neurologiques tels que myokémie, migraine, crises épileptiques ou chorée. Pour d'autres conditions, l'ataxie constitue plutôt des manifestations cliniques sans être le symptôme principal, par exemple la maladie d'Alexander, les maladies des prions héréditaires ou la neuroferritinopathie (van Gaalen and van de Warrenburg 2012).

Table II: Ataxies cérébelleuses dominantes

Name	OMIM #	Locus	Gene	Protein	Mutation	% of ADCA	Geographical distribution	Characteristic features
Repeat Expansions: coding								
SCA1	164400	6p22.3	<i>ATXN1</i>	Ataxin 1	CAG repeat	6–27%	Common: South Africa, Japan, India, Italy, Australia	Hyperreflexia, sensory neuropathy, mild cognitive impairment
SCA2	183090	12q24.12	<i>ATXN2</i>	Ataxin 2	CAG repeat	13–18%	Common: United States, Spain, India, Mexico, Italy	Polycytopathy, parkinsonism, dysphagia
SCA3	109150	14q32.12	<i>ATXN3</i>	Ataxin 3	CAG repeat	20–50%	Most common worldwide	Spasticity, polycytopathy, dystonia, parkinsonism
SCA6	183086	19p13.2	<i>CACNA1A</i>	Calcium channel, voltage dependent, P/Q type, α_{1A} subunit	CAG repeat	13–15%	Common: United States, Germany, Australia, Taiwan	Late onset, pure ataxia
SCA7	164500	3p14.1	<i>ATXN7</i>	Ataxin 7	CAG repeat	3–5%	Finland, Mexico, South Africa	Retinal degeneration
SCA17	607136	6q27	<i>TBP</i>	TATA box-binding protein	CAG repeat	Rare	United Kingdom, Belgium, France, Germany, Japan	Dementia
DRPLA	125370	12p13.31	<i>ATN1</i>	Atrophin 1	CAG repeat	0.8:100,000 (Japan) Rare worldwide	Japan, Portugal, United States	Dementia, epilepsy
HD	143100	4p16.3	<i>HTT</i>	Huntingtin	CAG repeat	3–7:100,000	Worldwide	Chorea, dementia
Repeat Expansions: noncoding								
SCA8	608768	13q21.33	<i>ATXN8OS</i>	Ataxin 8 opposite strand	CTG repeat	3%	Common: Finland	Pure ataxia
SCA10	603516	22q13.31	<i>ATXN10</i>	Ataxin 10	ATTCT repeat	Unknown	Mexico, Brazil	Seizures
SCA12	604326	5q32	<i>PPP2R2B</i>	Protein phosphatase 2, regulatory subunit B, β	CAG repeat	Rare worldwide 7% in India	Common: India	Tremor, polycytopathy
SCA31	117210	16q21	<i>BEAN</i>	Brain expressed associated with NEDD4	TGGAA repeat	8–40% in Japan Rare worldwide	Japan esp. Nagano prefecture	Spasmodic torticollis
SCA36	614153	20p13	<i>NOP56</i>	Nuclear protein 56	GGCCCTG repeat	6.3% Galicia, Spain 9 Japanese families	Spain, Japan	Motor neurone involvement
Conventional mutations								
SCA5	600224	11q13	<i>SPTBN2</i>	Beta 3 Spectrin	Deletions, missense mutations	Rare	United States, Germany, France	Pure ataxia, facial myokymia, gaze palsy
SCA11	611695	15q15.2	<i>TTBK2</i>	Tau tubulin kinase 2	Nonsense, frameshift deletions/insertions	Rare	United Kingdom, France, Germany	Pure ataxia
SCA13	605259	19q13.3-q13.4	<i>KCNK3</i>	Potassium channel, voltage-gated, shaw-related subfamily, member 3	Missense	1% (France)	France, Philippines	Early onset, mental retardation
SCA14	176980	19q13.4	<i>PRKCG</i>	Protein kinase C gamma	Missense, deletion	2% (France)	United Kingdom, France, Netherlands, United States, Japan, Australia	Myoclonus, dystonia
SCA15	606658	3p26-p25	<i>ITPR1</i>	Inositol 1, 4, 5-triphosphate receptor type 1	Multi-exon or whole gene deletion, missense	1.8% (France) 0.3% (Japan)	United Kingdom, France	Pure ataxia
SCA20	608687	11q12	–	–	260kb duplication	Rare	Australia	Dentate calcification, bulbar symptoms
SCA23	610245	20p13	<i>PDYN</i>	Prodynorphin	Missense	Rare	Netherlands	Pyramidal signs
SCA27	609307	13q33.1	<i>FGF14</i>	Fibroblast growth factor 14	Missense	Rare	Netherlands	Onset with tremor, psychiatric episodes
SCA28	610246	18p11.21	<i>AFG3L2</i>	ATPase FAMILY GENE 3-LIKE 2	Missense, deletion	3%	Italy, France, United Kingdom	Ophthalmoplegia, spasticity
SCA35	613908	20p13	<i>TGM6</i>	Transglutaminase 6	Missense	Rare	China	Pure ataxia

Hersherson J et al Hum Mut 2012 Sep; 33 (9): 1324-32

■ 1.1.3 Ataxies cérébelleuses à transmission liée au chromosome X

Il y a quelques formes d'ataxie cérébelleuse qui ont une transmission liée au chromosome X : SCAX1, SCAX2, SCAX3, SCAX4 and FXTAS. La plus commune est le syndrome du X-fragile associé à des tremblements et de l'ataxie (FXTAS). Le syndrome du X-fragile est causé par des mutations dans le gène *FMR1*. Il s'agit d'expansion CGG de plus de 200 répétitions (van Gaalen and van de Warrenburg 2012). Dans le cas de FXTAS, il s'agit d'expansion intermédiaire se situant entre 55 et 200 répétitions. Les individus affectés de FXTAS sont, majoritairement, des hommes de plus de 45 ans, des cas de femmes atteintes ont été reportés. La maladie est caractérisée par des tremblements, de l'ataxie cérébelleuse et des changements comportementaux. Parfois des signes parkinsoniens, une neuropathie périphérique, des problèmes cognitifs et des dysfonctions autonomiques sont aussi observés. Le groupe de maladie SCAX est un groupe de maladies hétérogènes cliniquement et génétiquement. SCAX1

est caractérisé par une hypotonie à la naissance, retard du développement moteur, ataxie à la marche, difficulté à garder l'équilibre, dysarthrie et mouvement des yeux lent (Bertini, des Portes et al. 2000). SCAX1 est causé par des mutations dans le gène *ATP2B3*. Aucun gène n'a été identifié pour les autres formes de SCAX.

▪ 1.1.4 Ataxies cérébelleuses à transmission mitochondriale

Les maladies mitochondriales sont causées par des mutations dans l'ADN mitochondriale ou bien par des mutations dans des gènes nucléaires ayant une fonction au niveau de la mitochondrie (Scharfe, Zaccaria et al. 1999). Habituellement, il y a une atteinte multi-système, qui inclue les systèmes nerveux périphérique et central, les yeux, le cœur et les glandes endocrines (van Gaalen and van de Warrenburg 2012). Le phénotype d'ataxie cérébelleuse est très commun dans les maladies mitochondriales.

Parmi ceux-ci, il y a la maladie mitochondriale causée par des mutations dans le gène *POLG*. *POLG* est un gène nucléaire codant pour une polymérase γ , et qui a pour fonction la réplication de l'ADN mitochondriale (van Gaalen and van de Warrenburg 2012). Les phénotypes associés à des mutations dans le gène *POLG* sont très variable : syndrome d'Alpert, l'ophtamoplégie progressive externe chronique, neuropathie, l'épilepsie, parkinsonisme et l'ataxie cérébelleuse. Les symptômes se manifestent habituellement après l'âge de 40 ans.

▪ 1.2 Approche pour identifier des loci candidats et des gènes causant des maladies héréditaires

Le séquençage du génome humain a permis de donner une position physique à presque tous les gènes. Par contre, pour beaucoup les fonctions demeurent inconnues. L'association de mutations pathologiques dans un gène peut aider grandement à la découverte de la fonction de

ce dernier (Carr, Markham et al. 2011). Bien que les techniques aient grandement évoluées dans les dernières années, une stratégie positionnelle reste dominante. Les ADN des patients soupçonnés d'avoir la même maladie sont collectés et ensuite comparé entre eux ou avec d'autres membres de leurs familles afin d'identifier des loci, c'est-à-dire des régions potentielles où pourraient résider le gène muté. Si une région candidate est partagée, les gènes qui y résident sont séquencés afin d'identifier des mutations.

▪ 1.2.1 Marqueurs génétiques

Les techniques de génotypage ont pour but de permettre l'identification de loci candidat par l'utilisation de marqueurs génétiques (microsatellites ou SNP). Ces techniques sont très importantes dans la recherche de gènes causant des maladies héréditaires.

Traditionnellement, les études de génotypage étaient effectuées avec des marqueurs microsatellites (STR). Les STR sont des marqueurs multialléliques et polymorphiques ce qui rend ce type de marqueurs très informatifs (Carr, Johnson et al. 2011). Les plateformes de marqueurs créés pour les analyses de génotypage ont habituellement une distance d'environ 10cM entre chaque marqueur (Carr, Diggle et al. 2012). L'amplification par une réaction de polymérase en chaîne (PCR) que nécessitent ces marqueurs peut être difficile à optimiser et peut causer des erreurs.

Depuis quelques années, l'utilisation des marqueurs microsatellites a été remplacée par les marqueurs de polymorphismes de simple nucléotide (SNP). Ces marqueurs étant biallélique, ils sont moins informatifs mais offrent une plus grande densité (Carr, Diggle et al. 2012). Cette plus grande densité qu'offrent les SNP permet d'obtenir une plus grande quantité d'information sur l'hérédité (Carr, Johnson et al. 2011).

▪ 1.2.2 Analyses génétique

L'utilisation de programmes informatiques est nécessaire pour l'analyse statistique des données obtenues par le génotypage microsatellite ou SNP. Plusieurs approches peuvent aussi être utilisées selon les données familiales disponibles et le mode de transmission de la maladie.

Les analyses de liaison classique nécessitent des informations détaillées sur le pedigree des familles à l'étude (Carr, Johnson et al. 2011). Ces analyses nécessitent de connaître la valeur des fréquences alléliques des marqueurs, ce qui peut être plus difficile à obtenir lorsque nous étudions une cohorte provenant d'une population plus homogène génétiquement. Les analyses de liaison peuvent être paramétriques, ce qui utilise la méthode statistique de localisation chromosomique qui repose sur le mode de transmission. L'analyse paramétrique calcule la probabilité qu'un ou plusieurs marqueurs sont liés à un trait génétique donné qui ségrège dans une famille (Sobel, Papp et al. 2002). Ceci correspond à un calcul « LOD score » multipoint lorsque l'analyse prend en compte des génotypes de plusieurs marqueurs voisins (par exemple les logiciels Genehunter et simwalk2) (Kruglyak, Daly et al. 1996) (Sobel, Papp et al. 2002). Le LOD score correspond à la probabilité que le ou les marqueurs ségrègent avec la maladie par hasard. Par convention, un LOD de 3 est considéré comme une association significative (1/1000 chance que ce soit le hasard) (Griffiths AJF 1993).

Une autre méthode d'analyse utilisée est la cartographie par homozygotie. Cette analyse recherche des régions homozygotes partagées entre des individus atteints d'une même famille ou de familles différentes atteintes de la même maladie. Cette méthode s'est avérée être très efficace pour cartographier des maladies récessives surtout en présence de familles consanguines (Carr, Markham et al. 2011). Cette méthode peut aussi être très performante lorsque les familles proviennent d'une population avec un effet fondateur. Dans ce cas, il est important de connaître la fréquence allélique des marqueurs et des haplotypes puisque des régions d'homozygotie pourraient être communes dans cette population par descendance commune sans être associées à la maladie (Sherman et al 2008). Des logiciels comme AutoSNPa permettent ce type d'analyse (Carr, Flintoff et al. 2006).

La méthode d'analyse par partage d'haplotype peut être très efficace lorsque nous sommes en présence de cas potentiellement hétérozygotes composés ou bien d'une maladie à transmission autosomique dominante. Un logiciel tel que Dominant Mapper ou Phaser permettent ce genre d'analyse, mais nécessite la présence d'au moins une paire de parents dans le pedigree (Carr, Johnson et al. 2011). Il est même possible de comparer plusieurs familles afin de voir si elles partagent l'un des haplotypes liés à la maladie.

▪ 1.2.3 Séquençage

Il n'y a pas seulement les techniques de génotypage qui ont beaucoup évoluées dans les dernières années. Les techniques de séquençage, essentielles à l'identification de mutations, ont évoluées à une très grande vitesse.

La méthode de séquençage traditionnelle, aussi appelé séquençage Sanger, a été inventée en 1977 par Frederick Sanger. Cette technique est basée sur une incorporation sélective par une ADN polymérase de didéoxynucléotides modifié avec un fluorochrome afin de terminer la réaction en chaîne (Sanger, Nicklen et al. 1977). Des fragments de toutes les tailles sont ainsi produits. Les produits migrent sur des capillaires et un laser effectue la lecture des fluorochromes dans un ordre correspondant à la taille des fragments. De cette façon, la séquence d'ADN du fragment désiré est obtenue. La limitation de cette technique est qu'un seul fragment est séquencé à la fois. Dans l'identification de mutations liées à une maladie, les exons d'un gène d'intérêt ainsi que les jonctions intron-exon sont amplifiées et séquencées en premier.

Bien que la méthode de séquençage Sanger soit toujours utilisée, elle a été surclassée par les technologies de séquençage de nouvelle génération (NGS). Cette révolution dans le monde du séquençage a débuté en 2005 et a rapidement transformé la recherche biomédicale. Ces nouvelles techniques nous permettent maintenant de séquencer la totalité des régions codantes d'un individu (séquençage de l'exome) et même la totalité du génome (séquençage du génome) (Xia, Wang et al. 2012). Ceci diminue de manière considérable le temps d'identification de

variants qui pourraient être pathogènes. Plusieurs plateformes et technologies de capture et séquençage ont été développées. Certains des désavantages de ces technologies sont la petite taille des séquences produites et un plus haut taux d'erreurs. Le séquençage traditionnel est toujours la méthode de choix pour la validation des variants potentiels identifiés par le séquençage NGS.

▪ 1.3 Projet de caractérisation génétique d'une forme d'ataxie tardive.

Nous avons recruté une cohorte de patients atteints d'une forme d'ataxie cérébelleuse à début tardif (Late Onset Cerebellar Ataxia, LOCA). Bien que la majorité des maladies tardives ont une transmission dominante et que les signes cliniques des patients LOCA sont compatible avec ce mode, les pedigrees et les histoires familiales sont suggestif d'un mode de transmission autosomique récessif. Deux grappes régionales sont représentées dans la cohorte: du Saguenay Lac-St-Jean et des Cantons de l'Est. Le but du projet de maîtrise résumé dans ce mémoire était d'identifier le gène responsable de cette ataxie tardive LOCA. Notre hypothèse est que les individus atteints de notre cohorte partagent des mutations dans le même gène et probablement la même mutation surtout si les patients sont originaires de la même région. Le premier objectif fut d'effectuer une analyse par génotypage afin d'identifier des loci candidats. Le second objectif fut d'utiliser les techniques NGS afin d'identifier des variants candidats.

Chapitre2 identification de Loci LOCA candidats

▪ 2.1 Introduction

On retrouve dans la population canadienne-française une certaine homogénéité génétique. Cette homogénéité est le fruit de plusieurs facteurs incluant un nombre limité de colons Français ayant contribué au pool génétique des canadiens-français (CF) (Laberge, Michaud et al. 2005). La Nouvelle-France a reçu plus de 30 000 Français, mais seulement de l'ordre de 10 000 ont laissé une descendance en Amérique, et c'est surtout 3 000 d'entre eux qui par leur arrivée précoce au 17^{ème} siècle ont le plus contribué au patrimoine génétique. Au début de la colonisation, les habitants se sont installés le long des rives du Fleuve St-Laurent. Après la conquête de la Grande-Bretagne, les CF ont continué à coloniser les régions plus éloignées du Fleuve (Laberge, Michaud et al. 2005). Les barrières de langues et religieuses et l'absence d'autres populations ont fait en sorte qu'il n'y ait pas eu de mélange interethnique.

Cette démographie génétique particulière fait en sorte que nous observons dans la population CF des maladies à effet fondateur (exemple: dystrophie musculaire oculopharyngée (OPMD) ou l'ataxie de Charlevoix-Saguenay (ARSACS)) (Laberge, Michaud et al. 2005). Inspiré d'autres travaux sur des maladies fondatrices, nous avons développé un projet de recherche de mutations historiques partagées qui pourrait expliquer la présence de cas familiaux d'ataxie cérébelleuse tardive (LOCA), plus particulièrement au Saguenay.

▪ 2.2 Résultats

▪ 2.2.1 Cohorte LOCA

Au moment de mon arrivée au laboratoire en 2010, la cohorte était composée de 38 individus atteints provenant de 17 familles (Figure 1). Nous avons regroupé les familles selon leur

région : 19 atteints originaires du Saguenay-Lac St-Jean (SLSJ), 6 atteints provenant des Cantons de l'Est et finalement 13 atteints d'origine régionale diverse (Figure 2).

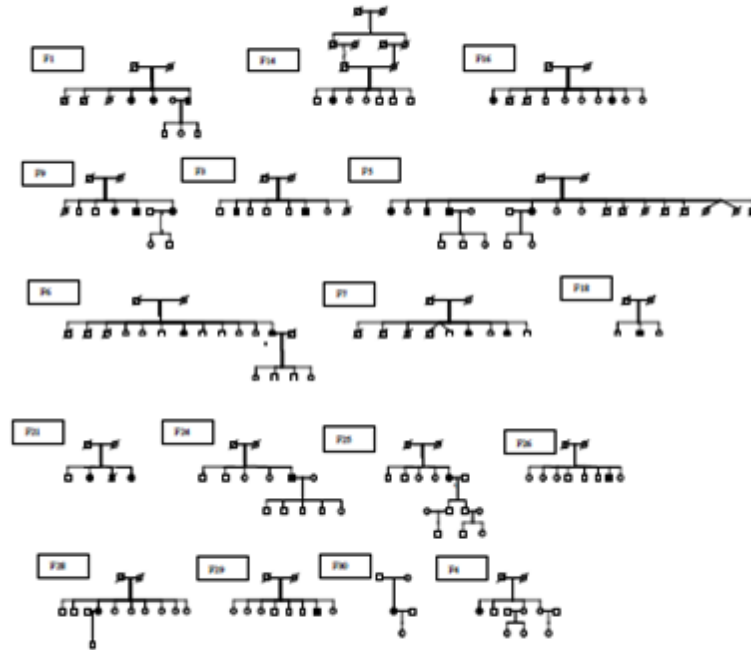


Figure 1: Pedigrees des familles LOCA.

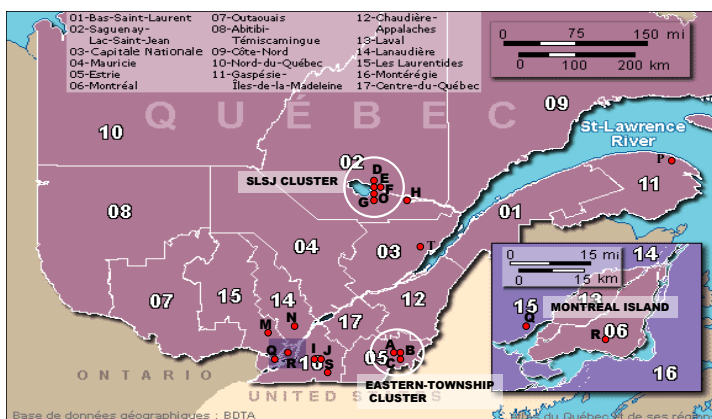


Figure 2: Distribution géographique des cas LOCA. Deux grappes de patients sont observées, l'une dans la région du SLSJ et une seconde dans les Cantons de l'Est.

Les principaux symptômes de LOCA sont une présence d'ataxie et de dysarthrie (88%) (Table III). Au niveau de l'imagerie on observe une atrophie cérébelleuse et une atrophie du lobe frontal dans la moitié des cas. Dans 52% des cas, un mouvement oculaires, appelé nystagmus, est observé. Les premiers symptômes sont observés tardivement, en moyenne à l'âge de 60 ans. Seulement 29% des participants avaient perdu la capacité de marcher au début de l'étude.

▪ 2.2.2 Exclusion du locus situé sur le chromosome 13

Avant mon arrivée au laboratoire, des études génétiques ont été effectuées afin d'identifier un locus candidat dans la cohorte LOCA. Les membres d'une famille provenant des Cantons de l'Est (2 atteints; famille 16) (Figure 1) ont été envoyés pour criblage entier du génome chez DeCODE genetics (Reykjavik, Islande). Ce criblage a été effectué avec 500 marqueurs microsatellites (STS) dispersés dans tout le génome. Un LOD score supérieur à 0,9 a été obtenu pour 9 loci (Figure 3). Le LOD le plus élevé a été observé sur le chromosome 13, les individus atteints de la famille étaient homozygotes pour deux marqueurs STS consécutifs. Pour tous les autres loci les individus atteints étaient hétérozygotes et ils semblaient y avoir un partage d'haplotype SNP plus commun.

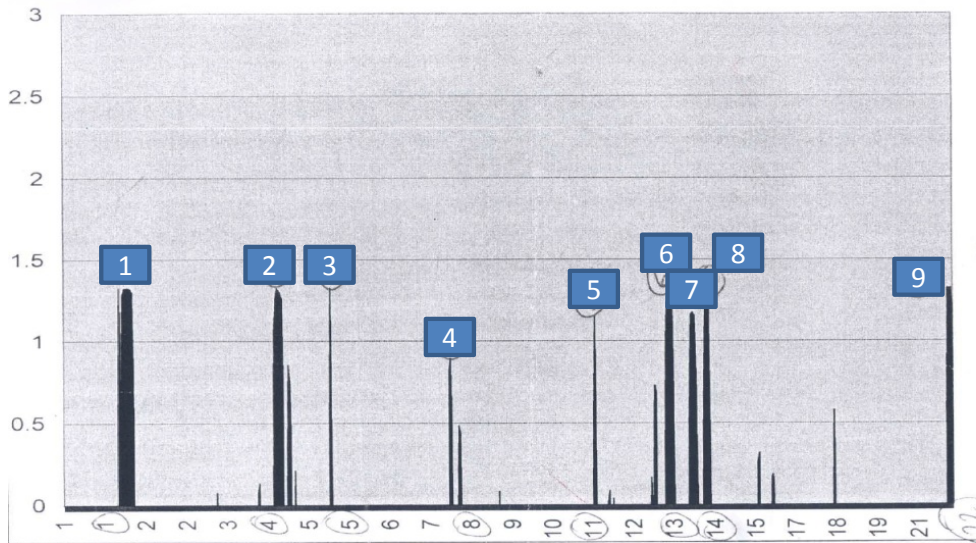


Figure 3: Graphique démontrant les résultats de LOD obtenu pour le criblage entier du génome par marqueurs STS.

Un criblage fin avec une plus grande densité de marqueurs STS dans la région du chromosome 13, suggérait une liaison génétique pour plusieurs familles. La région candidate ainsi déterminée était une région de 9.9cM (Figure 4). Une analyse de partage d'haplotype suggérait une diminution de la grandeur de l'intervalle candidat à 4cM. Un LOD score cumulatif pour plusieurs familles de près de 5 semblait aussi confirmer que nous avons trouvé le bon locus (Fig. 4).

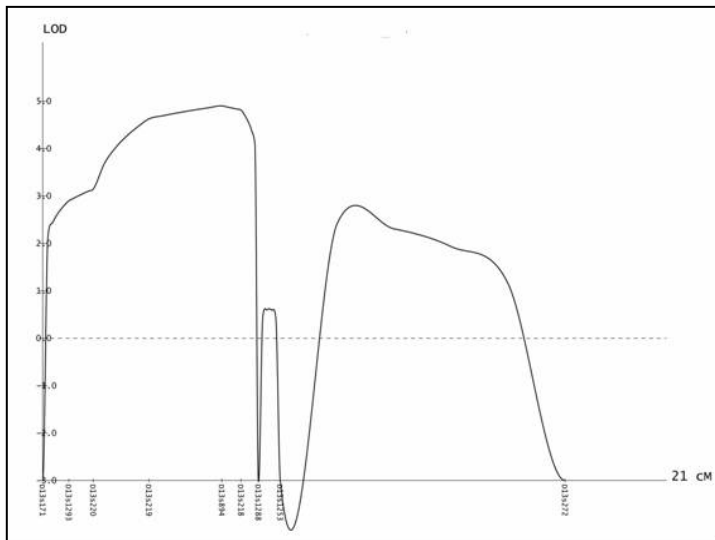


Figure 4: LOD score multipoint pour les familles LOCA dans la région candidate du chromosome 13. Un LOD maximal de près de 5 fut obtenu en additionnant le LOD de plusieurs familles.

Malgré un séquençage des presque tous les exons et jonctions exon/intron des gènes dans la région aucune mutation ne fut découverte. Nous avons donc décidé d'effectuer un autre criblage entier du génome cette fois en utilisant des marqueurs de simples nucléotides (SNP) (McGill Université et le Centre d'innovation Génome Québec). Cette stratégie permettait d'obtenir une couverture beaucoup plus dense de la région avec 700 000 marqueurs dispersés dans le génome (puce OmniExpress, Illumina). Une analyse des haplotypes dans nos familles les plus informatives a confirmé le partage sur le chromosome 13 par plusieurs cas LOCA d'un même haplotype de plus de 1Mb. Par contre, près de 60% des individus CF non-atteint de LOCA dans notre banque SNP partageaient aussi ce même haplotype de façon homozygote. Nous avons donc conclu que le partage d'haplotype observé dans la région du chromosome 13 chez nos malades, et même les résultats de la liaison cumulative, correspondaient au fait que les marqueurs dans la région sont peu informatives combiné à une plus grande homogénéité génétique chez les CF dans cette région. Ces analyses, en excluant le locus sur le chromosome 13, nous obligèrent à développer de nouvelles stratégies de recherche de locus candidats LOCA.

▪ 2.2.3 Analyse de partage d'homozygotie entre les familles LOCA

Le début de ma maîtrise a coïncidé avec le retour à la case départ. À ce moment, nous nous sommes tourné vers une analyse plus globale des données du criblage par marqueurs SNP afin d'identifier de nouveau loci candidat. En utilisant les données de génotypage déjà disponible au laboratoire et en reposant sur l'hypothèse d'une mutation fondatrice prépondérante, j'ai recherché des régions homozygotes partagées entre les individus LOCA. Pour se faire, j'ai utilisé les programmes Genome Studio (Applied Biosystem) et AutoSNPa afin d'identifier toute les régions homozygotes de plus de 1Mb (Carr, Flintoff et al. 2006). Ces deux logiciels permettent de visualiser les régions d'homozygotie et d'effectuer des comparaisons inter-individus afin d'extraire les régions partagées. Un partage d'homozygotie entre plus de deux patients LOCA de familles différentes a été observé pour 16 régions (Figure 5; Table IV). La région avec le plus de partage se trouve sur le chromosome 10 (région 10.1) d'une grandeur de 0.7Mb, l'homozygotie est présente chez 10 individus. La petite taille de cette région suggère un manque d'informativité des marqueurs plutôt qu'une association avec la maladie.

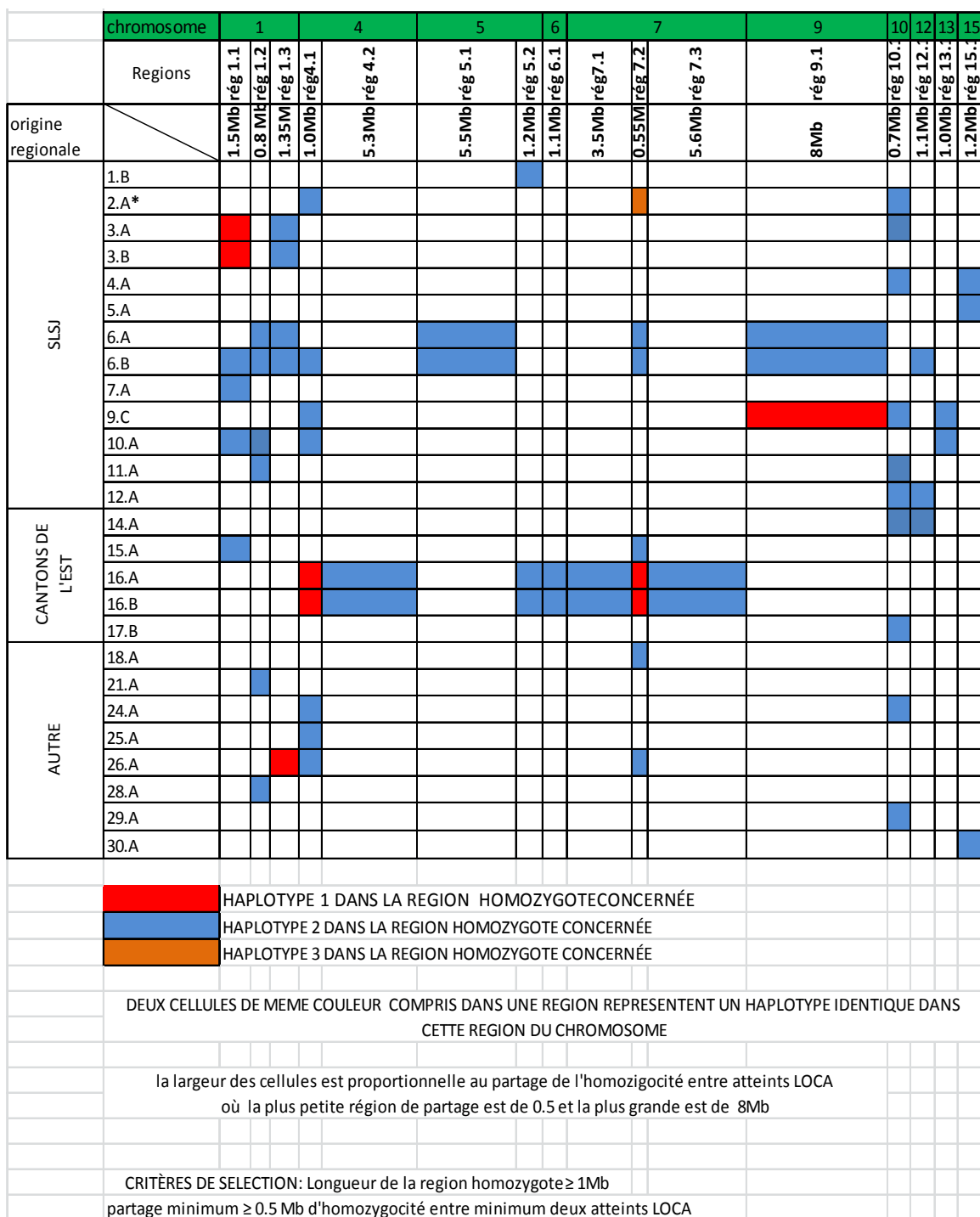


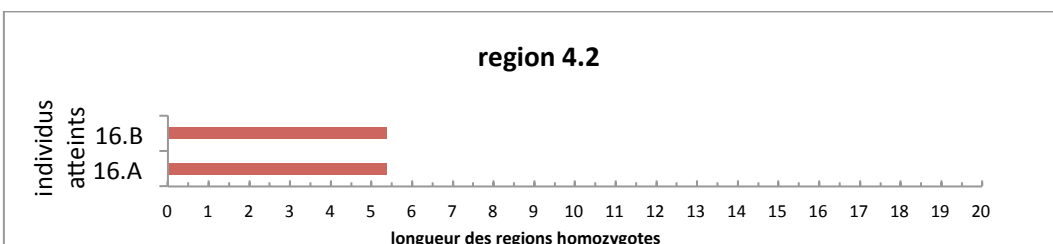
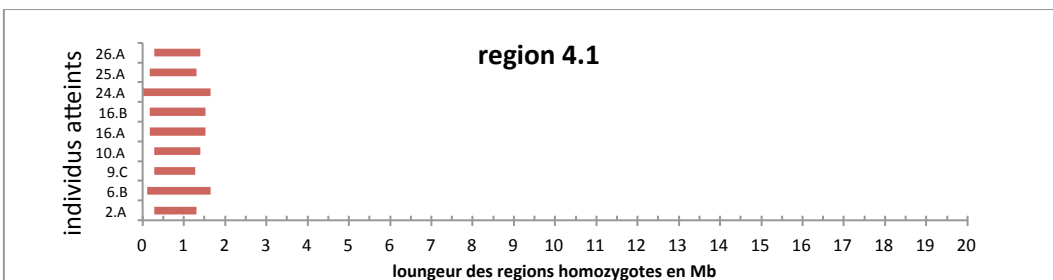
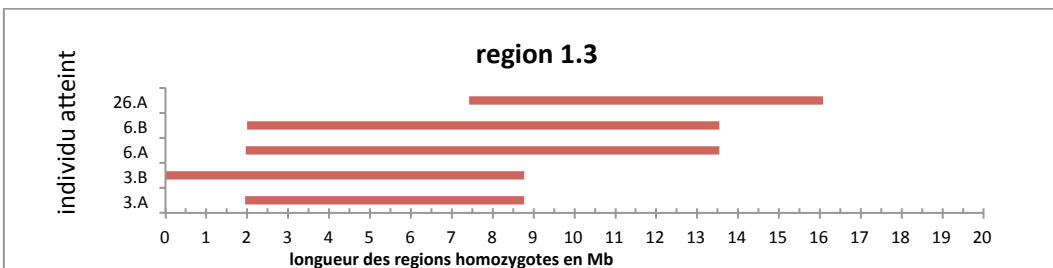
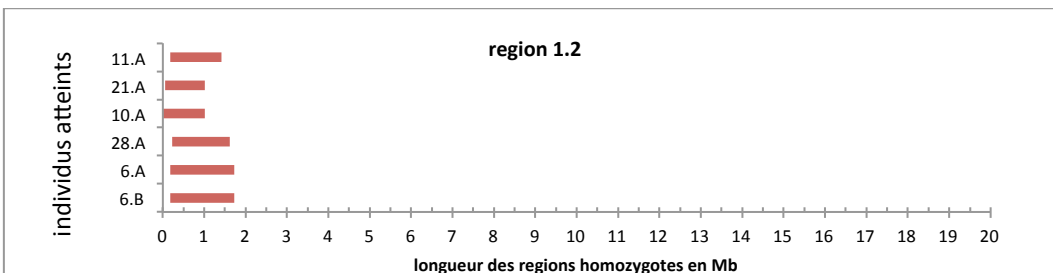
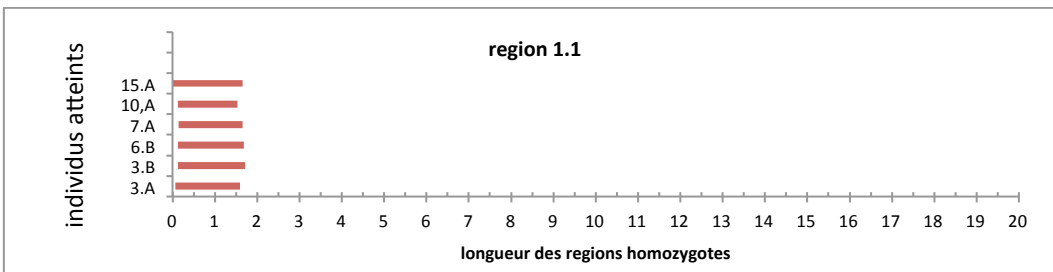
Figure 5: Analyse de partage d'homozygotie entre les atteints LOCA. La largeur des colonnes est proportionnelle à la taille de la région partagée. Les couleurs correspondent à des haplotypes différents.

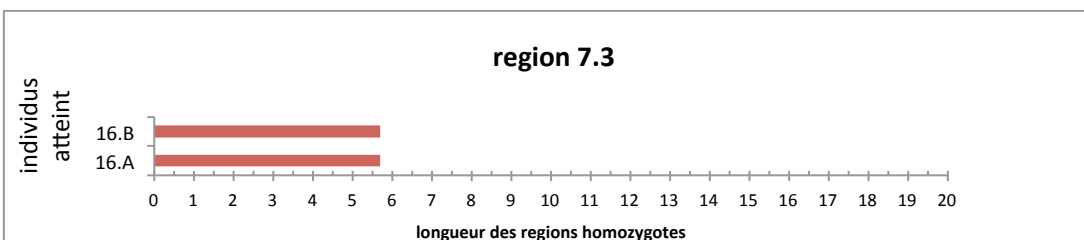
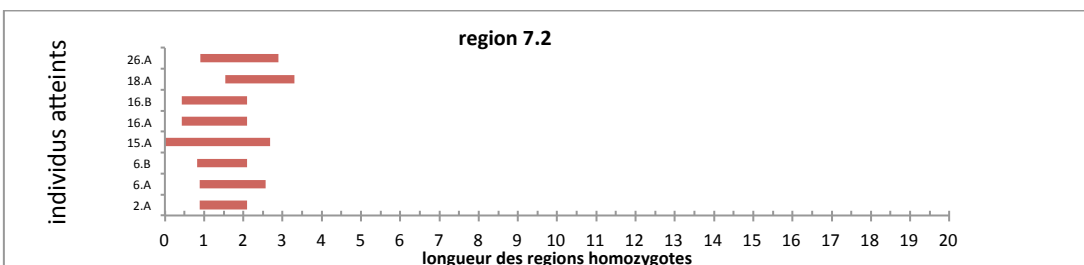
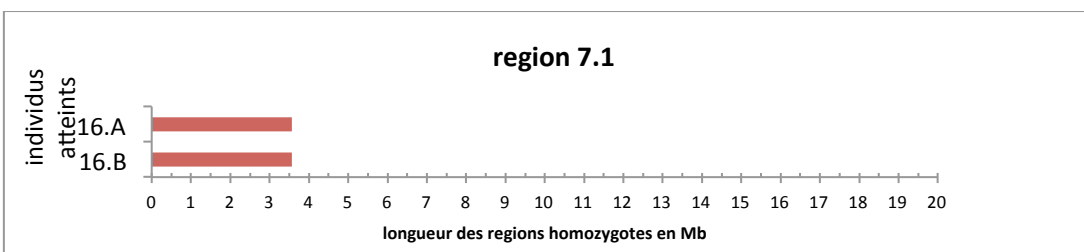
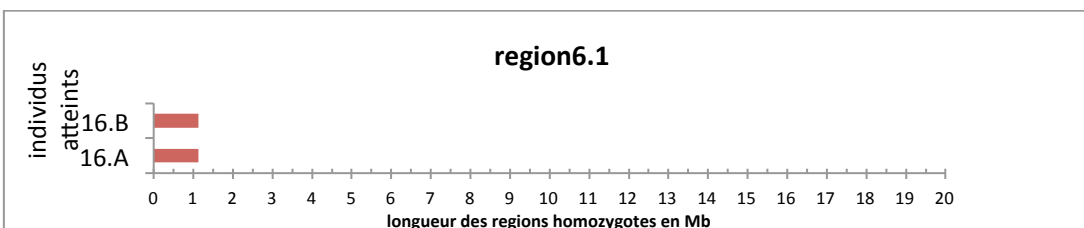
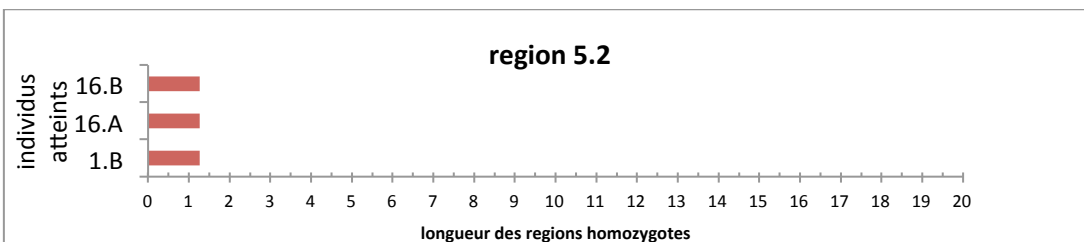
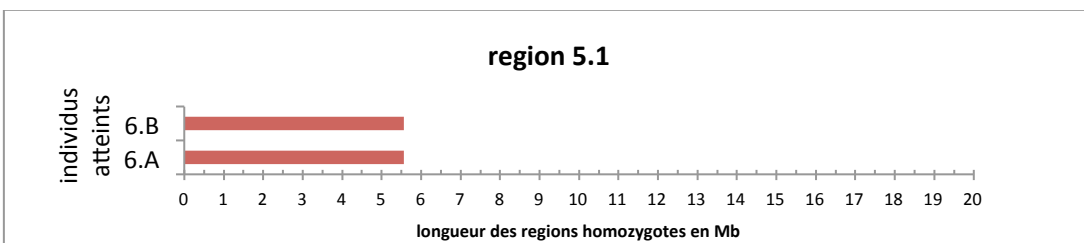
Table III: Résumé de l'analyse de partage d'homozygotie entre les atteints LOCA.

Chromosome 1				
region1.1		start	end	
	4	48 918 101	50 507 982	(1 532 045)
	5	48 975 596	50 507 982	(1 538 881)
	9	48 983 157	50 538 569	(1 555 412)
	10	48 995 491	50 507 982	(1 512 491)
	15	48 983 157	50 377 924	(1 394 767)
	20	48 849 243	50 507 982	(1 658 739)
region1.2				
	8	102 689 502	104 237 852	(1 548 350)
	9	102 689 502	104 237 852	(1 548 350)
	15	102 739 738	104 131 400	(1 391 662)
	16	102 515 063	103 535 627	(1 020 564)
	28	102 571 645	103 535 627	(963 982)
	35	102 689 502	103 930 141	(1 240 639)
region1.3	numero de l'attein	start	end	longueur
	4	202 341 926	209 160 213	(6 818 287)
	5	200 389 236	209 160 213	(8 770 977)
	8	202 349 561	213 930 201	(11 580 640)
	9	202 382 100	213 930 201	(11 548 101)
	33	207 836 526	216 864 203	(8 657 677)
chromosome4				
region4.1				
	3	33 432 805	34 441 990	(1 009 185)
	9	33 257 915	34 790 089	(1 532 174)
	14	33 432 805	34 415 180	(982 375)
	15	33 432 805	34 549 321	(1 116 516)
	21	33 317 239	34 666 752	(1 340 493)
	29	33 317 259	34 666 752	(1 340 493)
	31	33 143 198	34 795 406	(1 652 208)
	32	33 317 259	34 441 990	(1 124 731)
	33	33 432 805	34 549 321	(1 116 516)
region 4.2				
1803	21	48 082 608	53 446 689	(5 364 081)
1800	22	48 082 608	53 446 689	(5 364 081)
region 4.2				
1803	21	48 082 608	53 446 689	(5 364 081)
1800	22	48 082 608	53 446 689	(5 364 081)
chromosome5				
region5.1				
2461	8	10 089 432	15 651 629	(5 562 197)
2443	9	10 089 432	15 651 629	(5 562 197)
region5.2				
2390	2	97 997 353	99 257 197	(1 259 844)
1803	21	97 997 353	99 257 197	(1 259 844)
1800	22	97 997 353	99 257 197	(1 259 844)
chromosome6				
region6.1				
1803	21	46 152 196	47 274 142	(1 121 946)
1800	22	46 152 196	47 274 142	(1 121 946)
chromosome7				
region 7.1				
1800	21	63 020 558	66 592 283	(3 571 725)
1803	22	63 020 558	66 592 283	(3 571 725)
region 7.2				
2522	3	118 143 344	119 352 292	(1 208 948)
2461	8	118 143 344	119 634 840	(1 691 496)
2443	9	118 087 626	119 352 292	(1 264 666)
2910	20	117 258 785	119 943 543	(2 684 758)
1803	21	117 694 340	119 352 292	(1 657 952)
1800	22	117 694 340	119 352 292	(1 657 952)
1571	25	118 793 652	120 553 190	(1 759 538)
3015	33	118 154 347	120 159 363	(2 005 016)
region7.3				
1803	21	131 636 959	137 322 080	(5 685 121)
1800	22	131 636 959	137 322 080	(5 685 121)
chromosome9				
région9.1				
2461	8	13 667 557	21 673 289	(8 005 732)
2443	9	13 661 296	21 673 289	(8 011 993)
2387	14	8 990 033	23 593 120	(14 603 087)
chromosome10				
région10.1				
2522	3	73 400 079	75 363 938	(1 963 859)
1035	4	73 286 581	74 805 452	(1 518 871)
2770	6	74 032 361	75 371 861	(1 339 500)
2387	14	67 887 766	78 438 635	10 550 890
2739	16	73 610 497	75 054 988	(1 444 491)
1167	17	73 610 497	75 375 983	(1 765 486)
2316	19	73 507 426	75 231 922	1 724 496
2423	24	74 032 361	75 343 107	(1 310 746)
1245	31	73 659 190	75 363 938	(1 704 748)
2263	37	74 032 361	75 375 983	(1 343 622)
chromosome12				
région12.1				
2443	9	110 379 655	111 510 284	(1 130 629)
1167	17	110 238 980	111 524 326	(1 285 346)
2316	19	110 118 664	111 510 284	1 391 620
chromosome13				
région13.1				
	14	40 044 301	41 048 498	(1 004 197)
	15	40 008 884	41 048 498	(1 039 614)
chromosome15				
région 15.1				
	6	69 813 033	71 074 205	(1 261 172)
	7	69 813 033	71 074 205	(1 261 172)
	38	69 884 515	71 136 686	(1 252 171)

L'analyse du partage d'homozygotie démontre que la majorité des régions sont plutôt petites. En terme de grandeur des régions, la région qui semble la plus intéressante serait la région 1.3

sur le chromosome 1 (Figure 6). Dans cette région, partagée par trois familles, chaque individu est homozygote sur au moins 6Mb. La région commune aux trois familles est d'environ 1,35Mb.





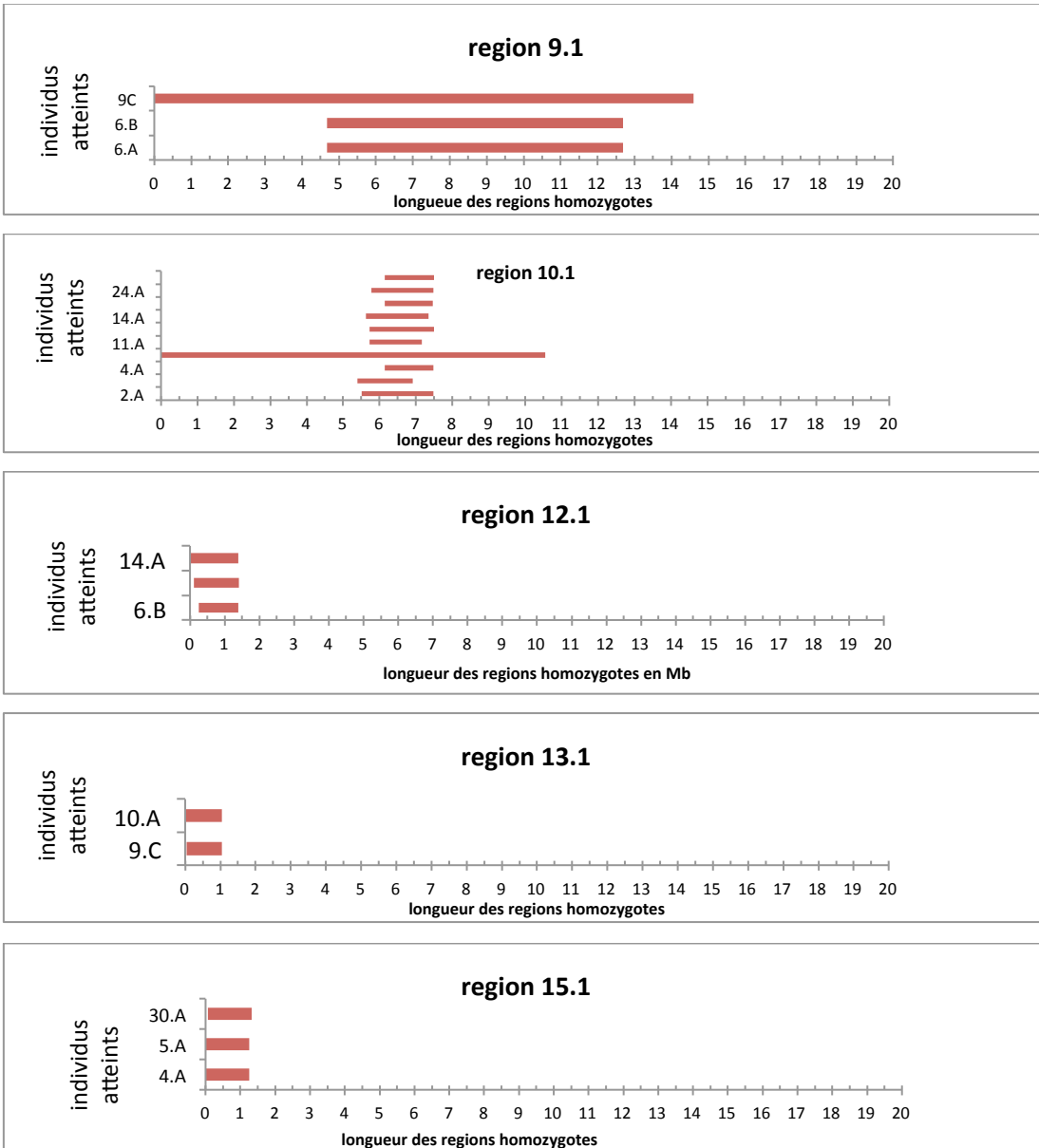


Figure 6: Partage d'homozygotie entre les atteints LOCA. Les graphiques démontrent les régions de plus de 0,5Mb partagée entre plus de 2 atteints LOCA.

Même en ayant effectué une liste de priorisation de ces régions d'homozygoties (Table IV), l'hypothèse de partage d'homozygotie ne nous a pas permis d'identifier avec certitude une région partagée entre les patients LOCA même ceux originaires d'une même région. J'ai résumé dans la table V les conclusions obtenues lors de l'analyse fine des régions identifiées par recherche de régions homozygotes partagées par au moins deux familles. Ces résultats

suggèrent qu'il existe peu de régions où une mutation historique transmise sur un haplotype ancestral pourrait expliquer une fraction des malades.

Table IV: Liste de priorités d'intérêt des régions d'homozygotie partagées entre les

priorité	régions	Raisons de la priorisation	conclusion
1	1.3	1.35 Mb de partage+ regroupe 5 atteints de 3 familles	région d'intérêt
2	7.2	0.55Mb de partage+ regroupe 8 atteints de 6 familles	région d'intérêt
3	9.1	8Mb de partage+regroupe 3 atteints de 2 familles	région d'intérêt
4	10.1	0.7Mb de partage + regroupe 10 atteints de 10 familles	région d'intérêt
5	1.2	0.9Mb de partage+regroupe 6 atteints de 5 familles	région d'intérêt
	4.1	0.8 Mb de partage+regroupe 9 atteints de 8familles	Exclusion : un frère atteint pas homozygote
	1.1	1.3Mb de partage+regroupe 6 atteint de 5 familles	Exclusion : un frère atteint pas homozygote
6	4.2	5.3Mb de partage, régions regroupe deux atteints d'une famille	exclusion : seulement deux atteints sont homozygotes dans la même famille dans chacune des régions
	5.1	5.6Mb de partage, région regroupe deux atteints d'une famille	
	6.1	1.1Mb de partage, région regroupe deux atteints d'une famille	
	7.1	3.5Mb de partage, région regroupe deux atteints d'une famille	
	7.3	5.6Mb de partage, région regroupe deux atteints d'une famille	
7	12.1	0.5Mb de partage+regroupe 3 atteints de 3 familles	Exclusion : un frère atteint n'est pas homozygote
	13.1	1Mb de partage+ regroupe 2 atteints de 2 familles	Exclusion : Région commune*
	15.1	1Mb de partage+ regroupe 3 atteints de 3 familles	Exclusion : Région Commune*

* : régions observées fréquemment comme homozygote dans une cohorte témoin de criblage de SNP non LOCA.

▪ 2.2.4 Partage d'homozygotie selon les différentes régions du Québec.

Puisque nous avons étudié le partage d'homozygotie chez les atteints LOCA, nous avons décidé d'utiliser ces données pour décrire l'homozygotie observée selon les différentes régions du Québec en espérant que ceci fasse ressortir qu'une région plus grande que la moyenne est

partagée entre cas de la même région. J'ai subdivisé les patients selon les principales grappes géographiques : SLSJ et Cantons de l'Est. Le nombre de régions homozygotes observées dans les différentes grappes est assez uniforme, c'est-à-dire que nous n'observons pas plus de régions d'homozygoties chez les individus du SLSJ comparativement au Cantons de l'Est ou autre région géographique du Québec (Figure 7).

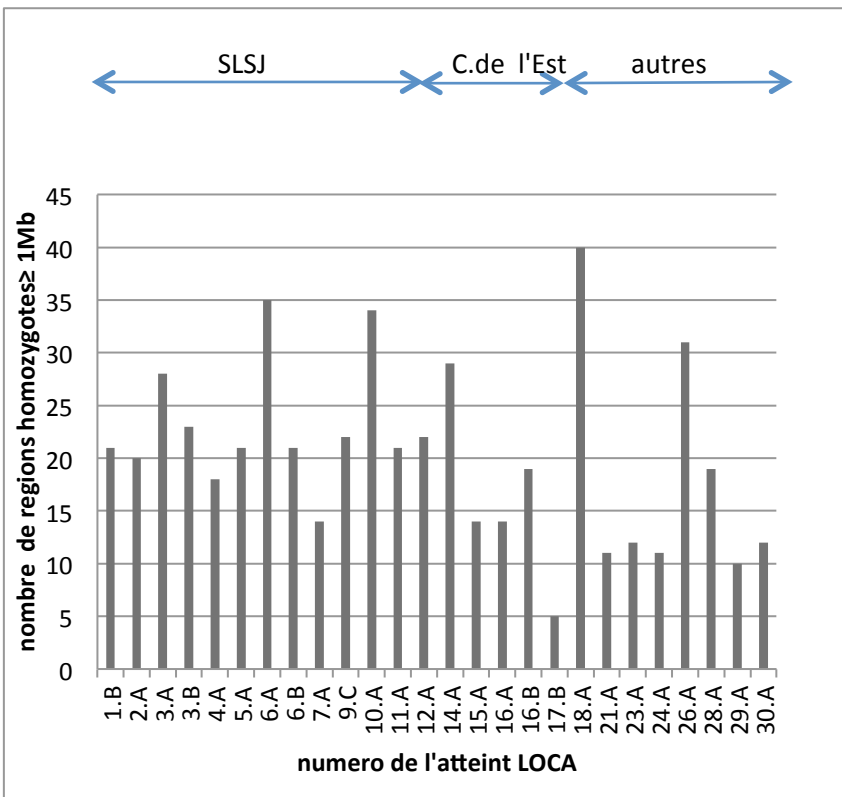


Figure 7 : Nombre de régions homozygote de plus 1Mb de longueur observés selon la région géographique.

Nous avons aussi regardé la grandeur de ces régions d'homozygoties afin de vérifier si dans certaines régions géographiques du Québec les régions homozygotes étaient plus grandes (Figure 8). Nous n'avons pas observé de corrélation entre la région géographique des individus et la grandeur des régions homozygotes. En fait, nous observons des régions de plus de 6Mb

pour des cas de toutes les régions. Afin d'identifier les régions d'homozygotie commune aux individus d'une même grappe géographique, j'ai utilisé le logiciel AutoSNPa.

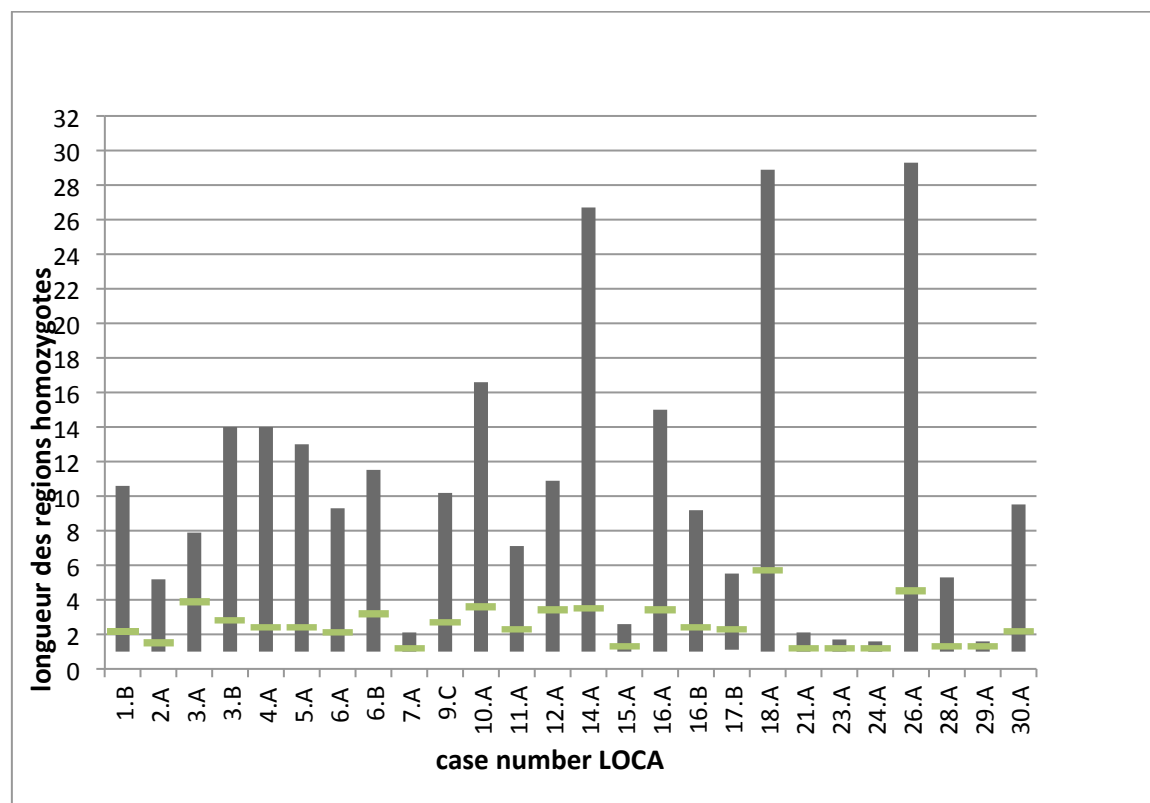


Figure 8: Longueur des régions homozygotes (minimum 1Mb) et longueur moyenne des régions homozygotes par atteint LOCA (barre de couleur verte).

▪ 2.2.5 Partage d'haplotype entre deux familles originaire des Cantons de l'Est.

À défaut d'avoir identifié des régions candidates par les approches précédentes, nous avons concentré nos efforts sur une famille consanguine originaire des Cantons de l'Est. J'ai comparé les régions homozygotes observées chez l'atteint de cette famille consanguine (famille 14) aux haplotypes observés chez les deux atteints (16A et 16B) de la famille 16, aussi originaire des Cantons de l'Est. L'analyse d'homozygotie chez l'individu 14A a révélé 29 régions

homozygotes de plus de 1Mb. Parmi ces 29 régions, un partage d'haplotype avec les atteints de la famille 16 a été observé pour 10 régions (Table VI). Les régions communes entre les deux familles étaient de petite taille, avec seulement trois régions de plus de 1Mb.

Table V: Partage d'haplotypes entre deux familles originaire des Cantons de l'Est.

chromosome.intervalle	longeurs de la region homozygoteoù se trouve le partage d'haplotype	coordonnées de l'haplotype commun	longueur de l'haplotype commun
2.1	26 746 586	188469784-189721391	1 251 607
2.2		192131348-192683449	552 101
2.3		192758287-193844067	1 085 780
5.1	11 935 828	86287327-87131861	844 534
6.1	1 948 507	121944568-122499397	554 829
7.1	1 265 333	71518642-72679066	1 160 424
10	1 724 496	73507426-74478051	970 625
15.1	7 593 746	31895202-32575858	680 656
15.2	1 094 016	69708893-70425766	716 873
16.1	1 104 507	21628515-22620480	991 965

▪ 2.2.6 Exclusion du gène candidat *SYT14*.

Au cours de ma maîtrise, un nouveau gène associé à une forme tardive d'ataxie récessive fut identifié (Doi, Yoshida et al. 2011). Une famille d'origine Japonaise composée de deux individus atteints présentant un retard psychomoteur et une ataxie spinocérébelleuse (SCA) a été associée au gène synaptogamin 14 (*SYT14*). La protéine SYT14 se lie aux phospholipides et pourrait avoir un rôle important dans le transport vésiculaire. Le pedigree de cette famille Japonaise suggère un mode de transmission autosomique récessif.

Une mutation faux-sens a été identifiée chez les deux atteints japonais changeant une glycine pour un acide aspartique (Gly484Asp) (Doi, Yoshida et al. 2011). Les auteurs suggèrent que cette mutation conduit à une perturbation de la machinerie d'exocytose, ce qui est un mécanisme reconnu de neurodégénérescence. SYT14 est principalement exprimé au niveau de

la rétine et du cervelet où il maintient l'intégrité des synapses (Doi, Yoshida et al. 2011). Dû au chevauchement clinique entre les cas japonais et notre cohorte CF, nous avons décidé de séquencer le gène afin de vérifier pour la présence de mutation dans notre cohorte.

Le séquençage des sept exons et des jonctions exon-intron de *SYT14* dans notre cohorte CF n'a pas révélé de mutation en ségrégation avec le phénotype d'ataxie. Les seuls variants identifiés sont communs et présent dans les bases de données internationales (Table VII).

Table VI: Variants identifiés dans SYT14 dans la cohorte CF.

Gène	fonction	expression	SNP	Variant
SYT14	Intégrité des synapses	Rétine cervelet	Rs2307890	c.765_677delAGA
			Rs2494187	c.1164+80A>T
			Rs17188183	c.1362T>C

2.3 Conclusion

En conclusion, les différentes approches utilisées reposant sur l'homozygotie partagée afin d'identifier des loci candidats dans notre cohorte LOCA CF ont démontré qu'aucunes régions homozygotes n'étaient partagée par un grand nombre des familles. Certaines des méthodes d'analyse que j'ai utilisé ont permis d'identifier des régions partagées, mais aucune par leur taille ou nombre de cas porteurs ne semblait mériter un passage au séquençage des gènes dans la région. Dû à l'homogénéité génétique plus grande observée chez les CF, il est devenu plus claire au cours de mon projet de maîtrise qu'il est très important de comparer les régions homozygotes avec une cohorte CF non atteints de la pathologie d'intérêt afin de ne pas s'attarder à un locus tout simplement commun chez les CF par descendance commune ou non-informativité des marqueurs comme celui du chromosome 13. Ceci étant dit, l'identification de régions partagées même entre un groupe plus limité de familles si combiné aux nouvelles technologies de séquençage, comme le séquençage exomique, pourraient aider l'identification

de variants d'intérêt. Dans le chapitre 3 je décris nos premiers efforts d'utilisation de cette approche croisée.

- **2.4 Matériels et méthodes**

- **2.4.1 Recrutement des patients**

Pour cette étude, nous avons recruté 38 individus atteints de LOCA (17 familles) ainsi que les membres non-atteints de ces familles lorsque cela nous était possible. Tous les participants ont signé un consentement de recherche approuvé par le comité d'éthique du Centre de recherche du centre hospitalier de l'Université de Montréal (CRCHUM). Les individus atteints ont été évalués par un ou des neurologues spécialisés dans les syndromes ataxiques. L'ADN des participants a été obtenu à partir de sang ou de salive. L'extraction a été effectuée selon la procédure standard en utilisant le kit Gentra (Qiagen) pour le sang et Oragene pour la salive.

- **2.4.2 Génotypage et analyses génétiques avec marqueurs STS.**

L'ADN des individus de la famille 16 originaire des Cantons de l'Est a été envoyé pour un criblage entier du génome à DeCode genetics (Reykjavik, Islande). Le criblage consistait dans le génotype de 500 marqueurs avec une distance moyenne de 9cM entre chaque marqueur. Les analyses de liaison paramétrique ont été effectuées avec le logiciel GeneHunter.

- **2.4.3 Génotypage et analyses génétiques avec marqueurs SNP.**

L'ADN des individus de la cohorte LOCA a été envoyé pour un criblage entier du génome au Centre d'innovation Génome Québec et Université McGill (Montréal, Canada). La puce OmniExpress (Illumina, USA) contenant plus de 700 000 marqueurs a été utilisée. Les analyses d'homozygoties ont été effectuées avec le logiciel AutoSNPa (Carr, Flintoff et al. 2006) et Genome studio (Illumina, USA).

▪ 2.4.4 PCR et séquençage du gène *SYT14*.

Les oligonucléotides (Table VIII) utilisés pour l'amplification des fragments de *SYT14* ont été conçus en utilisant le logiciel ExonPrimer (<http://genome.ucsc.edu>) selon la séquence de référence Hg19, afin de couvrir la totalité de la région codante ainsi que les jonctions exon-intron est synthétisée chez Invitrogen. La réaction PCR a été effectuée en utilisant 20ng d'ADN dans une réaction total de 8uL contenant 1X tampon PCR, 3nM MgCl₂, 10uM du mélange d'amorce sens et anti-sens et 0.4 unité de Taq Qiagen. La réussite du PCR a été vérifiée sur un gel d'agarose 1%. Les réactions de séquençage ont été effectuées au Centre d'innovation Génome Québec et Université McGill (Montréal, Canada) sur un appareil ABI3730 (Applied Biosystems, USA). L'Analyse des séquences a été faite avec le logiciel Seqman v.4.03 (DNASar, USA). Les variants identifiés ont été comparés à la base de donnée dbSNP afin de vérifier s'ils avaient déjà été rapportés et à quelle fréquence.

Table VII : Amorces utilisés pour l'amplification du gène SYT14.

EXON	AMORCE SENS	AMORCE ANTISENS	TAILLE (bp)
1	CGGGAAGGTTCCATTTCTTC	CAGACCACAGGAGGCGG	330
2	TTCTTTGTAACAGTCAGATTCA TGC	AAATACAAATTATCCGCCATC C	208
3	TTTCTGTGAAATGCCTTTATCC	GGCAGACAAACAGATCAATG G	408
4	TGAATGAATGCTGCTGAATG	GACAATTCACCTTTATGTAAC AGC	575
5	CAGTAGGGCCAGGTGAGG	TTCCTAGAATTCAGGCACAA G	360
6	TTGTAGACTATACTGTGTGTTG AGGC	ACCAAGAGGAGAAGCACAGG	410
7	GGCAGAATTGTTTTTCACATCTT C	AAAATAGACTGACAACTTG TTGGC	660
8	ACCTAATCCAAATAAACATGG C	AACTCCCTAATACTGAGAGTC CC	347
9	ACCCAACGATTGACATCTGG	ACTGCTGGCAAGTTTTAGGG	190

2.5 Contribution

J'ai effectué la totalité des analyses comparatives d'homozygotie chez les individus de la cohorte LOCA. Pour ce faire, j'ai développé des habilités d'analyse bioinformatique en utilisant le logiciel AutoSNPa. J'ai participé à l'extraction d'ADN. J'ai aussi effectué la construction des amorces, l'amplification ainsi que l'analyse des séquences avec le logiciel DNASStar.

Chapitre 3: Recherche de mutations par séquençage exomique.

▪ 3.1 Introduction

Les techniques de séquençage permettant l'identification des mutations ont grandement évoluées dans les dernières années. Le séquençage traditionnel (type Sanger) permet de séquencer des fragments sélectionnés, mais un fragment à la fois. Cette approche est bien adaptée au séquençage des gènes candidats localisés dans une région liée à une maladie héréditaire (Bras, Guerreiro et al. 2012). Malgré les désavantages du clonage positionnel qui reposait sur le séquençage de type Sanger, cette stratégie a mené à l'identification de centaine de gènes responsables de maladies neurologiques. Les nouvelles technologies de séquençage, dit de « Next Generation », permettent un séquençage rapide et peu dispendieux. Le séquençage exomique permet de séquencer la totalité des régions codantes du génome. L'obtention d'une aussi grande quantité de données génétiques a complètement modifié l'approche utilisée pour l'identification de mutations. Par contre, cette nouvelle technologie a aussi ses limites car elle ne couvre pas tous les gènes et ne permet pas de séquencer les régions introniques ou intergéniques. Le séquençage du génome entier permet de palier à la couverture intronique des gènes, mais tout comme le séquençage exomique la couverture est parfois incomplète et l'interprétation des variants introniques est plus difficile. Afin de nous aider à identifier le gène responsable de LOCA, nous avons opté pour le séquençage exomique.

▪ 3.2 Résultats

▪ 3.2.1 Séquençage exomique des patients LOCA

Six atteints ont été sélectionnés pour le séquençage exomique. Parmi les six atteints LOCA, 4 (deux familles) sont issus du SLSJ (1A, 1B, 9A et 9B), le 5ème d'une famille consanguine de Repentigny (13A), et le 6ème des Cantons de l'Est 16A (Figure 9 et 10). Pour la capture des régions exomiques nous avons utilisé la technologie SureSelect de Agilent (50Mb). La

couverture de la capture correspond à tous les exons annotés du projet GENCODE (ENCODE), tous les exons annotés dans la région codante (CDS) consensus (mars 2009, Hg19) et la séquence de petits ARN non-codant (miRbase et Rfam). Pour le séquençage, c'est la plateforme Illumina HiSeq2000 qui a été choisie. Le séquençage de nouvelle génération, tel que le séquençage exomique, produit une quantité énorme de données soit environ 10Gb par échantillon. L'analyse d'une telle quantité de données nécessite une infrastructure informatique adéquate avec des ordinateurs et serveurs très puissants. Comme nous n'avions pas directement accès à une telle infrastructure, le partenariat avec Perkin Elmer nous permettait d'accéder à un tel support informatique via un portail web. Nous étions donc en mesure d'effectuer nous-même les analyses de nos exomes et d'utiliser les différents outils bioinformatiques en ce connectant au serveur de Perkin Elmer. Suite à l'alignement des séquences avec la séquence de référence et la détection des variants (Li and Durbin 2009), les différents variants sont annotés (McKenna, Hanna et al. 2010). Cette annotation détermine l'emplacement du variant (exon, intron, région intergénique), l'effet fonctionnel du variant (faux-sens, nonsens, synonyme, jonction d'épissage) et si le variant a déjà été rapporté dans une base de donnée publique, dans notre cas dbSNP132. Nous obtenons donc une liste de tout les variants observés pour un individu donné, qui contient environ 20000 variants dans les régions codantes seulement (exclusion des variants d'épissage et des régions régulatrices introniques) (Stitzel, Kiezun et al. 2011). Les outils disponibles sur le serveur de Perkin Elmer nous permettent aussi de combinés les listes de variants annotés de plusieurs individus afin d'effectuer des comparaisons. Afin de déterminer le ou les variants candidats, nous devons finalement appliquer des filtres afin de réduire le plus possible la liste de variants. Avant d'effectuer des comparaisons inter-individus et d'appliquer notre hypothèse de recherche, j'ai effectué une recherche de variants rare dans les gènes connus comme causant une ataxie afin de ne pas introduire du bruit dans notre analyse comparative. Pour ce faire, j'ai créé un filtre recherchant les variants dans les gènes suivants : SPTBN2, SYNE1, GRM1, FXN, AFG3L2, CABC1, ADCK3, ATCAY, SETX, MRE11A, ZNF592, ANO10, TTPA, PPTX, TDP1, VLDLR, CA8, ATP8A2, ATM et WDR81. Aucun variant rare et prédit comme pathogène n'a été observé chez les six atteints séquencés.

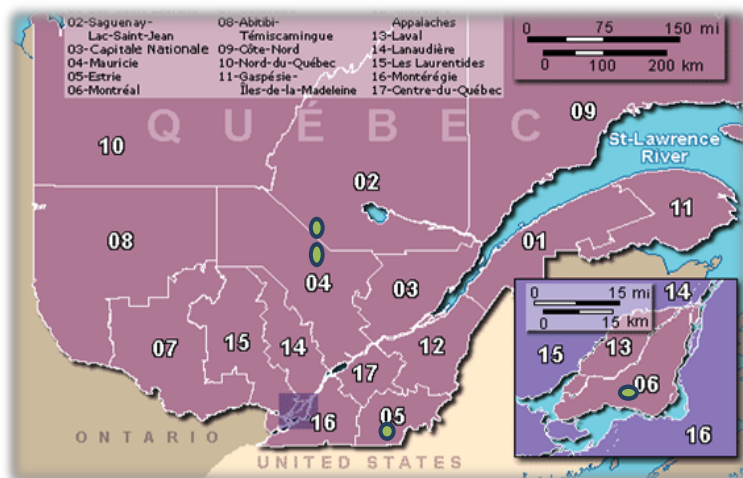


Figure 9: Distribution géographique des atteints LOCA analysé par séquençage exomique.

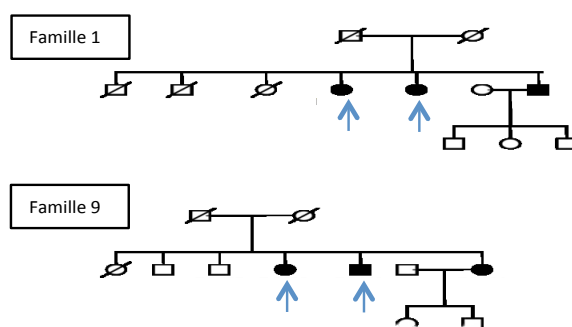


Fig10 : pedigrees des familles originaires du SLSJ envoyées au séquençage exomique

Suite à l'exclusion des gènes connus et selon notre hypothèse qui consiste en un mode de transmission récessif de la maladie, j'ai cherché des variants homozygotes ou hétérozygotes composés. Afin d'obtenir des variants candidats j'ai établi les critères de sélection suivants: 1) bonne qualité de couverture (au moins couvert 3 fois), 2) génotypes identiques chez les atteints d'une même famille, 3) non décrit dans dbSNP132, 4) le variant homozygote ou l'un des variants hétérozygotes composés devaient se trouver dans une séquence codante dont l'effet était non-synonyme (faux-sens ou nonsense) ou à une jonction d'épissage. Les sections qui suivent décrivent les différentes stratégies de comparaison utilisées afin d'identifier le gène responsable de LOCA.

▪ 3.2.2 Partage de variants entre les six atteints LOCA.

Étant donné la démographie génétique du Québec (Laberge, Michaud et al. 2005), nous avons tout d'abord comparé les six atteints LOCA ensemble à la recherche de variants partagés. Notre hypothèse était que tous les atteints partageaient au moins un même variant soit à l'état homozygote ou hétérozygote. Pour ce faire, j'ai donc combiné les listes de variants annotés des six atteints LOCA. En utilisant les filtres mentionnés dans la section précédente, nous avons identifié 355 variants partagés par les six atteints LOCA (Figure 11). Dû à ce grand partage, nous étions dans l'impossibilité d'identifier un gène causal. Nous avons donc établi de nouvelles stratégies d'analyse des données d'exomes.

▪ 3.2.3 Partage de variants dans les régions candidates identifiées par criblage entier du génome.

Comme première stratégie, nous avons consacré nos efforts dans les régions candidates qui avaient été identifiées par criblage entier du génome dans la famille 14. J'ai appliqué les mêmes filtres que dans la section précédente, mais cette fois afin de maximiser les chances d'un partage de variants j'ai combiné les listes de variants annotés des deux familles originaires du SLSJ afin de concentrer l'analyse sur ces deux familles. Cette analyse a permis

l'identification de 794 variants absents de la base de données dbSNP132, présent dans les régions codantes et avec ségrégation dans les familles. Aucun de ces variants n'était présent dans les loci candidats identifiées par analyse de liaison de la famille 14. Cette analyse n'a donc pas permis l'identification de variants candidats. (Figure 12).

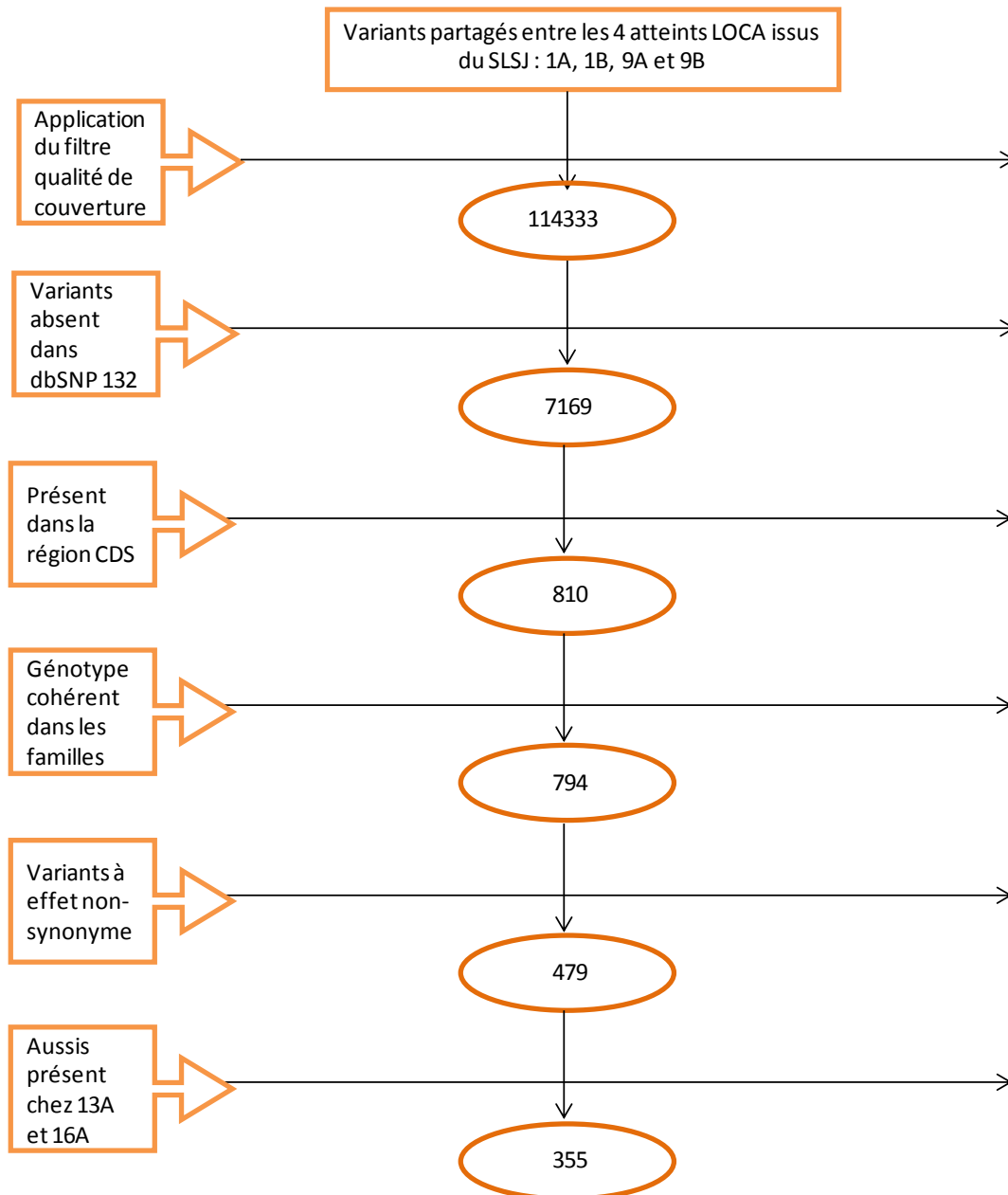


Figure 11: Représentation schématique de l'analyse partage de variants entre les six atteints LOCA.

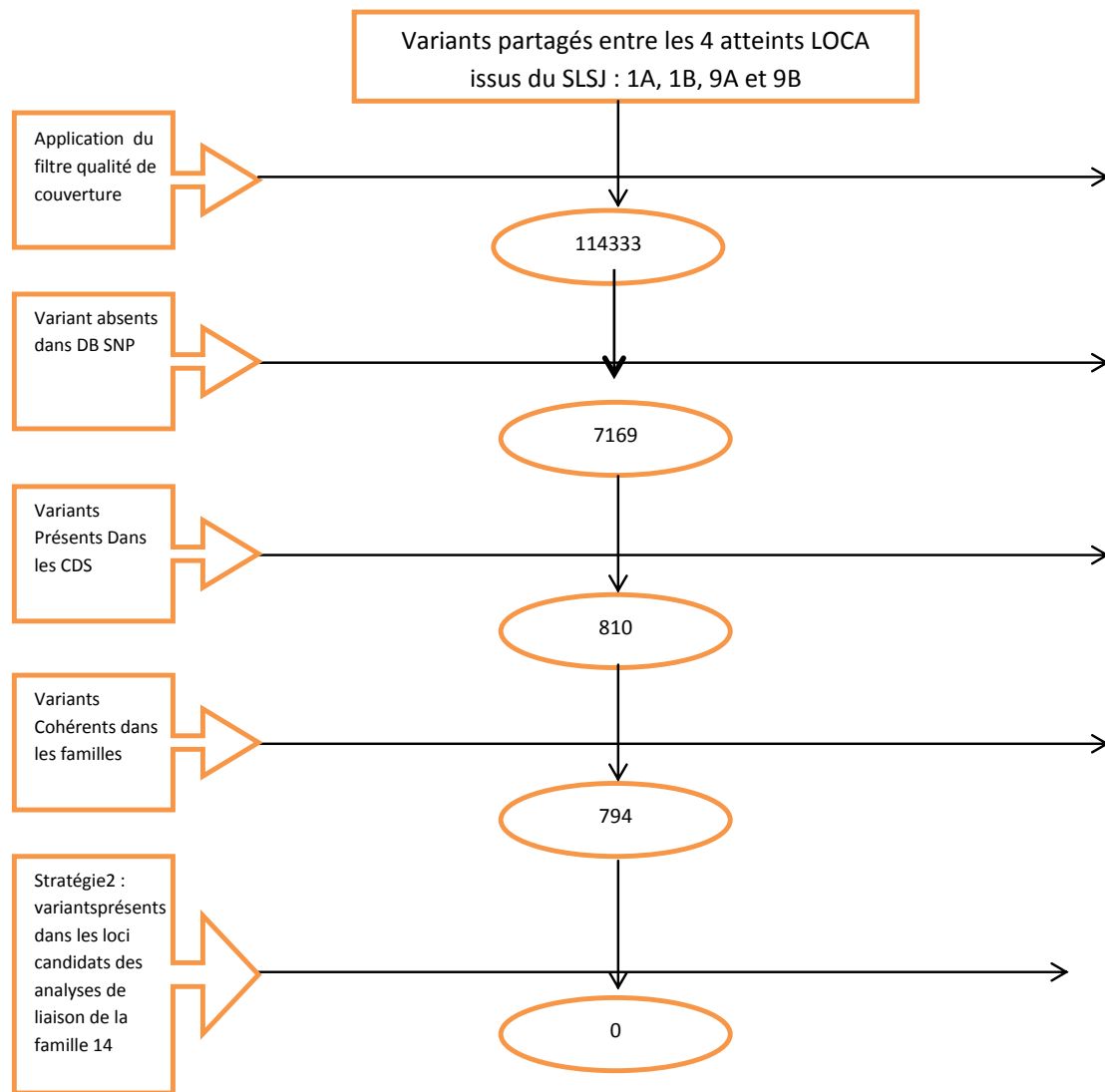


Figure 12: Répartition des variants après séquençage exomique dans les loci candidats identifiés par analyse de liaisons de la famille 14.

▪ 3.2.4 Partage de variants dans les régions candidates identifié par partage d'homozygotie.

Comme la stratégie précédente n'a pas fonctionné, j'ai décidé de regarder les variants présent dans les régions identifiées par partage d'haplotype dans le chapitre 2. Encore une fois, nous

avons utilisé les mêmes filtres et comparé les listes de variants annotés des quatre patients provenant du SLSJ. Encore une fois, aucun des 794 variants correspondant aux critères de sélection n'était présent dans les régions d'homozygotie. Cette analyse n'a pas permis d'identifier des variants candidats (Figure 13).

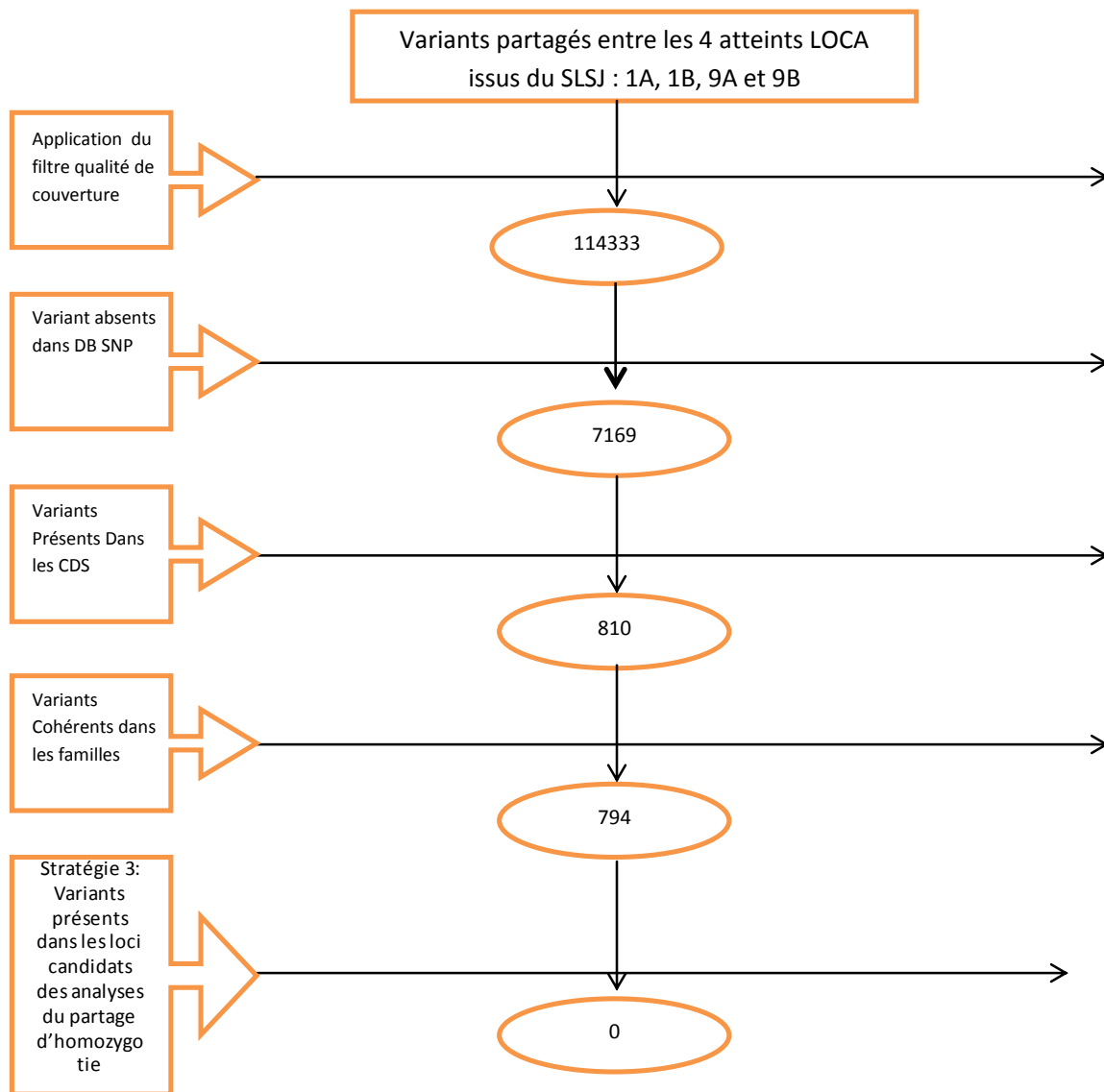


Figure 13: Répartition des variants après séquençage exomique dans les loci candidats identifiés par partage d'homozygotie entre les atteints LOCA.

▪ 3.2.5 Partage de variants dans les régions candidates identifiées par partage d'haplotype entre deux familles des Cantons de l'Est.

Dans le chapitre 2, nous avons identifié des régions de partage d'haplotype entre une famille consanguine originaire des Cantons de l'Est et une seconde famille provenant de la même région. Nous avons sélectionné les régions où la famille consanguine était homozygote. J'ai donc analysé les données de séquençage exomique des quatre atteints du SLSJ dans ces régions. Aucun des 794 variants correspondant aux critères de sélection n'était présent dans ces régions candidates. Cette approche n'a identifié n'a pas permis l'identification de variants candidats (Figure 14).

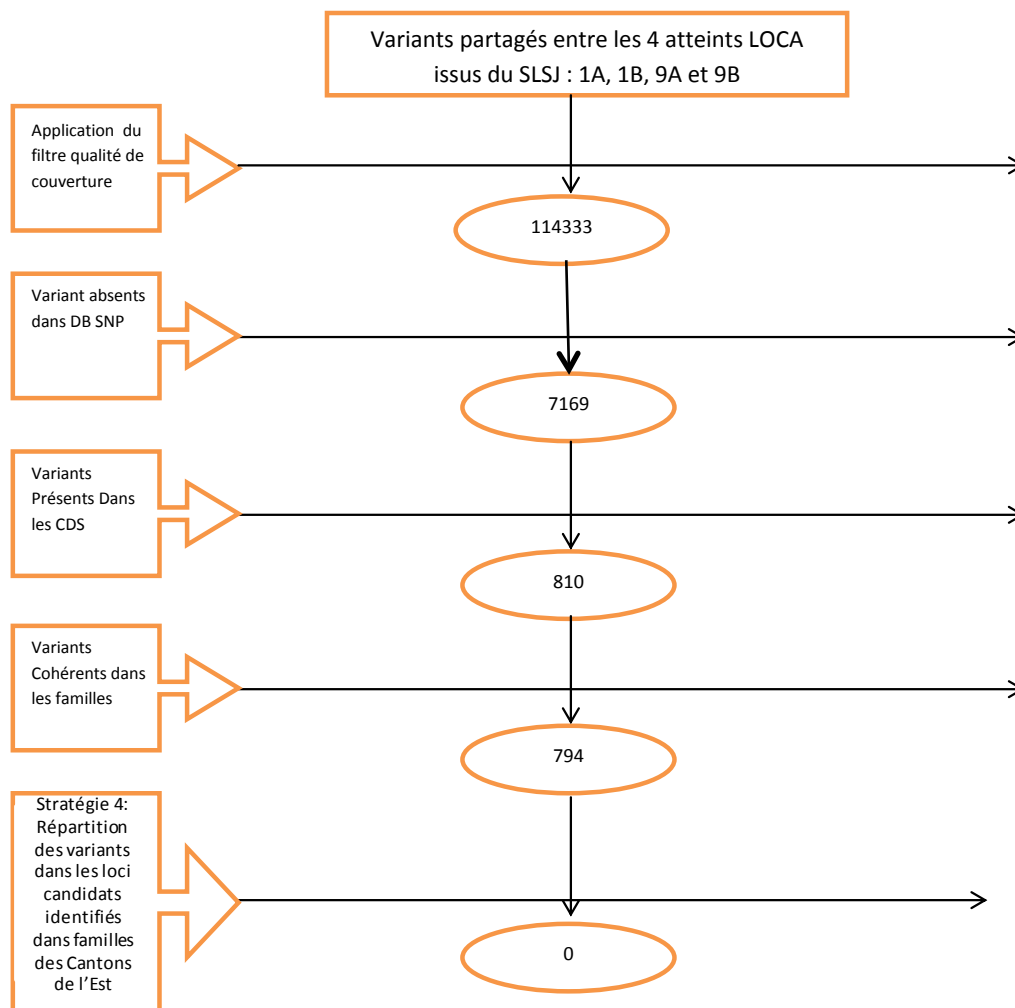


Figure 14: Répartition des variants après séquençage exomique dans les loci candidats identifié par partage d'haplotype dans deux familles originaire des Cantons de l'Est.

▪ 3.2.6 Partage de variants entre les quatre atteints LOCA originaire du SLSJ.

Après l'obtention de ces résultats nous avons réévalué notre approche d'analyse. Nous avons décidé de chercher un variant hétérozygote partagé entre les individus du SLSJ. Cette approche permettait de vérifier deux hypothèses. Premièrement transmission récessive de la maladie hétérozygote composé dont l'une des mutations est commune et donc partagée entre les atteints su SLSJ. Une fois le variant partagé identifié, une recherche d'un second variant dans ce même gène pour chacune des familles sera effectuée. Deuxièmement, comme la majorité des maladies tardives ont une transmission dominante, la liste des variants pourra être regardée sous cet angle. Nous avons comparé les quatre atteints provenant du SLSJ. Nous avons aussi introduit dans l'analyse trois individus âgés originaire du SLSJ mais non-atteint de LOCA. J'ai donc combiné les listes de variants annotés des patients LOCA avec celles d'individus contrôles afin de filtrer les variants fréquents dans cette région. Pour cette analyse nous avons utilisé les critères suivants: 1) variants hétérozygotes, 2) partagés par les quatre saguenéens (1A, 1B, 9A et 9B), 3) une bonne qualité de séquençage et de couverture, 4) génotype identique entre les atteints d'une même famille, 5) non décrit dans dbSNP132 et 1000 genomes, 6) variants dans une région codante du gène et ayant un effet non-synonyme ou un variant d'épissage, et 7) absent chez les trois individus contrôles saguenéens. Cette analyse nous a permis d'obtenir une liste de 50 variants candidats (Figure 15). Comme la liste obtenue est tout de même longue, j'ai confronté la liste aux données d'exome des deux autres patients (13A et 16A). Un total de 24 variants étaient aussi partagés avec 13A, un patient originaire de la région de Montréal et 16 variants partagés avec 16A originaire des Cantons de l'Est. En regardant les variants partagés par les six atteints LOCA, nous avons réduit la liste des variants candidats à quatre. Les quatre gènes candidats sont: mucin 14 (*MUC14*), low density lipoprotein receptor-related protein1B (*LRP1B*), COBW domain containing 1 (*CBWD1*) et zinc finger protein 585B (*ZNF585B*). Une validation de la ségrégation dans les familles ainsi qu'une validation fonctionnelle seront nécessaire afin de confirmer la pathogénicité de ces variants.

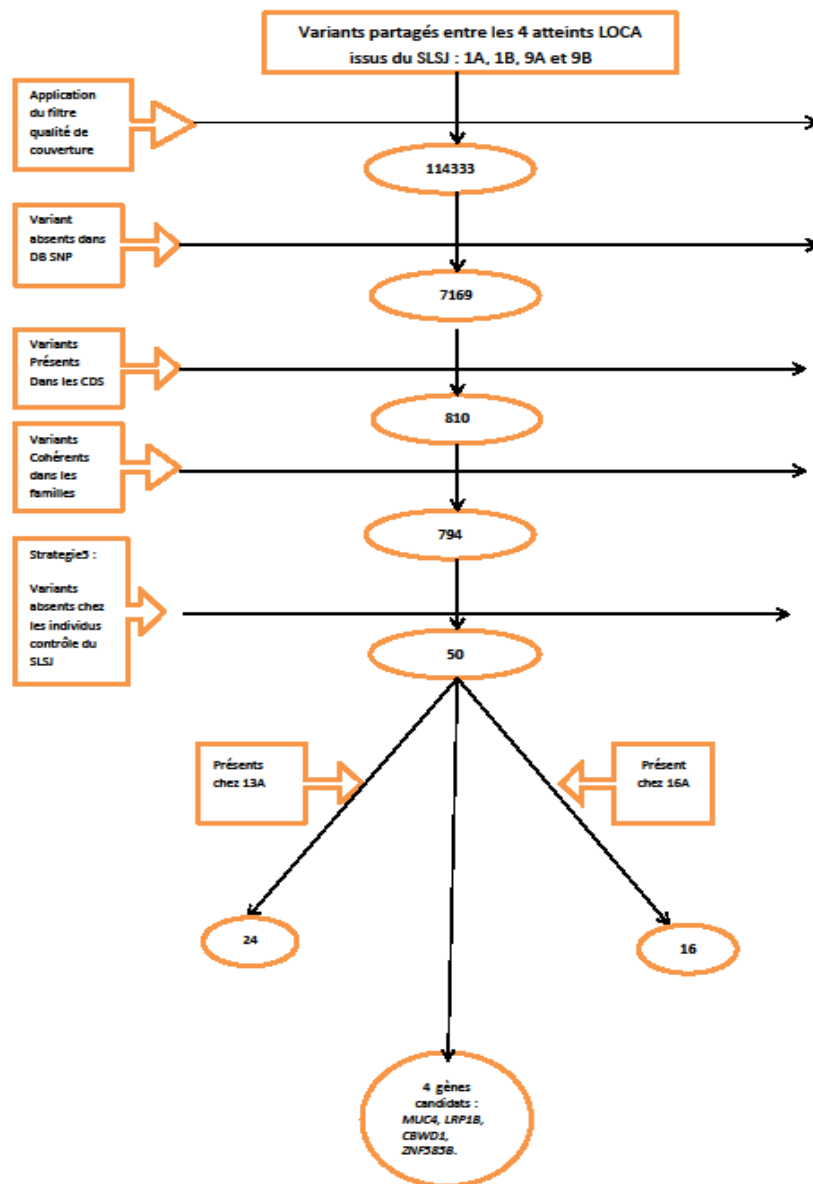


Figure 15: Répartition des variants après séquençage exomique entre les atteints LOCA selon un mode de transmission autosomique dominant.

▪ 3.3 Conclusion

Nous avons utilisé différentes approches afin d'identifier le gène responsable de LOCA. Le fait de combiner les données de génotype SNP des différentes analyses décrites au chapitre 2 avec le séquençage exomique ne nous a pas permis d'identifier des mutations récessives partagées entre les atteints LOCA. Ayant exclu le partage d'homozygotie entre les atteints LOCA, nous avons opté pour une hypothèse de mutation hétérozygote. Cette analyse nous a permis de significativement réduire la liste des gènes candidats, avec seulement quatre gènes partagés entre les six patients LOCA. L'analyse des données brutes de chacun de ces quatre gènes dans les familles n'a pas mené à l'identification d'un deuxième variant. Il faudrait vérifier la couverture de chacun de ces gènes afin de s'assurer qu'elle soit complète. Dans le cas où des exons seraient manquant, il serait important de compléter le séquençage manuellement. Les variants identifiés dans les quatre gènes devraient être séquencés afin de vérifier la ségrégation dans les familles et aussi pour la présence de ces mêmes variants dans les autres cas de notre cohorte LOCA ainsi que chez des contrôles. Ces dernières étapes permettront peut-être d'identifier le gène responsable de LOCA.

▪ 3.4 Matériels et méthodes

▪ 3.4.1 Séquençage exomique

Le séquençage exomique a été fait dans les laboratoires de Perkin Elmer (USA). Un totale de 5ug d'ADN génomique (concentration 100ng/uL) a été envoyée par individu. La capture des exons a été effectuée avec le kit SureSelect 50Mb (Agilent, USA) selon le protocole du manufacturier. Le séquençage a été effectué sur un appareil HiSeq 2000 (Illumina, USA). Des fragments de 100bp ont été séquencés par la méthode à extrémité jumelée (paired-end).

▪ 3.4.2 Analyse des données d'exomes

Les résultats des séquençages exomiques ont été analysés avec le logiciel GeneSifter (Geospiza, USA) via un portail web avec Perkin Elmer. L'alignement des séquences avec une séquence de référence (Hg19) a été effectué avec l'algorithme Burrows-Wheeler Aligner (BWA) (Li et al 2010) et l'annotation des variants avec GATK v.2.9 (Broad Institute). Les étapes de filtration ainsi que la visualisation des variants a été effectuée avec le logiciel GeneSifer.

3.5 Contribution

J'ai participé à la préparation des échantillons d'ADN pour le séquençage exomique. J'ai aussi effectué la totalité des analyses bioinformatiques (alignement, annotation, filtration, interprétation) sur les données de séquençage exomique.

Chapitre 4: Conclusions et perspectives

▪ 4.1 Conclusions du projet

Mon projet de maîtrise consistait en l'identification du gène responsable d'une ataxie cérébelleuse à début tardif (LOCA). À cette fin, une cohorte comportant 38 individus atteints provenant de 17 familles a été recrutée. Nous avons été en mesure de regrouper les patients selon deux régions géographiques importantes soit le Saguenay Lac-St-Jean (19 atteints), les Cantons de l'Est (6 atteints) et origine diverse (13 atteints). Afin d'atteindre l'objectif principal de ce projet de recherche, nous avons combiné des approches de regroupement géographique, de génotypage par SNP et de séquençage Sanger et exomique. Malgré les différentes hypothèses d'analyse utilisées, nous n'avons pas été en mesure d'identifier le gène causal de LOCA.

Même si une majorité des syndromes à début tardif ont une transmission dominante, nous avons opté pour un mode de transmission récessif de la maladie comme hypothèse de départ pour nos analyses. Le fait que les parents n'avaient pas d'histoire d'ataxie et le nombre d'atteint chez les enfants était suggestif d'un tel mode de transmission. Il est tout de même important de souligner, que compte tenu du début tardif des symptômes, les parents non pu être évalués en clinique étant déjà décédés lors du recrutement. De plus, en considérant l'âge de début des symptômes (en moyenne 60 ans) et l'espérance de vie à cette époque (71 ans en 1960, www.data.worldbank.org) il est très probable que les parents soient décédés avant l'apparition des signes cliniques. De plus, les connaissances médicales de cette époque n'étaient pas ce qu'elles sont maintenant, les symptômes apparents ont peut être été ignorés. L'apparition de trouble à la marche à un âge tardif est souvent considérée comme bénigne et rarement une investigation plus poussée n'est envisagée. L'absence de phénotype chez les parents est donc seulement basée sur l'histoire rapportée par les membres de la famille et non une évaluation rigoureuse.

Afin d'identifier un locus candidat, une approche de génotypage par marqueurs microsatellites et de marqueurs SNP a été effectuée. Une analyse traditionnelle de liaison n'a pas permis l'identification d'un locus candidat. La population CF est connue pour avoir un taux d'homozygosité plus élevé en comparaison à d'autres populations (Casals, Hodgkinson et al. 2013), nous avons donc analysé les SNP par cartographie d'homozygosité. Un partage d'homozygotie a été observé entre certains atteints, mais aucun loci ne semblait lier à la maladie avec confiance. Comme les ataxies sont des maladies hétérogènes au niveau clinique et génétique, notre incapacité à identifier un locus candidat par les différentes analyses de partage d'haplotype pourrait être due à la présence d'hétérogénéité génétique dans notre cohorte. C'est-à-dire que malgré la similarité phénotypique, la maladie pourrait être causée par différents gènes selon les familles. Nous avons beaucoup misé sur le regroupement régional des patients puisque plusieurs effets fondateurs ont été décrits chez les CF (Laberge, Michaud et al. 2005) (Bernard, Thiffault et al. 2010) (Tetreault, Srour et al. 2011) en particulier dans la région du SLSJ. Il est donc possible que malgré l'homogénéité clinique des patients LOCA, notre cohorte présente une plus grande hétérogénéité génétique que ce que nous avons prévu.

Comme l'analyse des données de séquençage exomique dans les régions d'homozygoties partagées n'a pas mené à l'identification de variants candidats nous avons décidé d'exclure la possibilité d'une mutation homozygote partagée et de changer d'hypothèse d'approche. Convaincu qu'un partage de variants devait exister entre les individus du SLSJ, nous avons décidé de chercher un variant hétérozygote partagé entre les individus du SLSJ. Cette approche permettait de vérifier deux hypothèses : 1) transmission récessive de la maladie hétérozygote composé dont l'une des mutations est commune et donc partagée entre les atteints du SLSJ. 2) transmission sous un mode autosomique dominant. Avec l'approche d'un variant hétérozygote partagé, nous avons obtenu une liste de 50 variants candidats, que nous avons été en mesure de réduire à quatre variants lorsque comparés avec les deux autres patients LOCA originaire de Montréal et des Cantons de l'Est. Des quatre candidats restant, *LRPIB* nous semble le gène le plus intéressant. Ce gène fait partie de la famille des récepteurs de lipoprotéines de faible densité (Liu, Li et al. 2001). Cette protéine de surface a la capacité de lier des ligands qui mène par la suite à un processus d'endocytose. Ce gène nous semble

comme un bon candidat puisqu'un autre gène de cette famille a déjà été associé avec un syndrome ataxique. Il s'agit de *VLDLR* codant aussi pour un récepteur de lipoprotéine de très faible densité. Tout comme LRP1B, le récepteur VLDLR effectue aussi l'internalisation de ces ligands dans un processus d'endocytose. Des mutations dans le gène *VLDLR* sont associées à une ataxie cérébelleuse avec retard mental (Boycott, Flavelle et al. 2005).

Beaucoup de travail reste à faire afin d'identifier le gène ou les gènes responsables de LOCA. Il faudrait tout d'abord vérifier la couverture des quatre gènes candidats, surtout *LRP1B*, afin de s'assurer qu'un deuxième variant n'ait pas été manqué. Dans le cas où des exons seraient manquants, il faudrait compléter avec le séquençage Sanger. L'accès à une grande cohorte de patients et aussi une grande quantité de données génétiques permettra certainement la découverte du ou des gènes responsables de cette maladie.

▪ 4.2 Perspectives

Pour la suite du projet LOCA, je crois que les quatre variants candidats identifiés lors de la dernière analyse devraient être validés au niveau génétique. Nous devrions regarder la ségrégation de chacun de ces quatre variants dans les familles et aussi regarder la présence de ces variants dans nos autres familles LOCA. Dans l'éventualité où un de ces variants ségrège parfaitement avec la maladie dans nos familles, nous pourrions envisager une validation fonctionnelle du variant, c'est-à-dire vérifier l'effet du variant sur la fonction de la protéine codée par le gène. Le type d'expérience effectuée à ce moment dépend directement de la fonction du gène dans lequel se trouve le variant.

Nous pourrions aussi envisager d'envoyer un plus grand nombre d'individus atteints de LOCA au séquençage exomique. Comme nous avons déjà regroupé nos familles en grappes régionales et que deux familles originaires du SLSJ ont déjà été séquencées, nous pourrions

nous concentrer sur les atteints de cette même région. Dans ce cas, l'analyse des données de séquençage s'intéresserait aux variants partagés par la totalité ou la majorité des atteints LOCA originaire du SLSJ. Compte tenu de l'homogénéité génétique présente dans une région comme le SLSJ, il sera important de comparer la liste de variants obtenus pour nos atteints LOCA avec les variants présent chez des individus non-atteints de LOCA aussi originaires du SLSJ et pour lesquels nous avons des données de séquençage. De cette manière, les variants fréquents dans cette sous-population mais non liés à la maladie pourront être éliminés. Comme il s'agit d'une maladie tardive, il est très important de ne pas être trop exigeant dans les étapes de filtrage contre des bases de données contrôles. Il est possible que des atteints soient présent dans ces bases de données mais ne présentaient aucun symptôme au moment du séquençage dû à leur jeune âge.

Il est possible que les limitations des technologies, tel que le séquençage exomique, pourraient être un obstacle à l'identification du gène causal de LOCA. Bien que le séquençage exomique permet de séquencer la grande majorité des exons et certaines régions régulatrice non-codante en une seule étape, certaines régions sont sous représentées. Comme il s'agit de séquençage exomiques, les introns ne sont donc pas couverts par cette expérience. De plus, les régions riches en contenu GC et les régions répétées ont un haut taux d'échec au séquençage (Gilissen, Hoischen et al. 2011). Plusieurs ataxies cérébelleuses, plus particulièrement les ataxies spinocérébelleuses à transmission dominante, sont causées par une expansion d'un triplet répété (van Gaalen and van de Warrenburg 2012). Ce type de mutation peut difficilement être détecté avec le séquençage de nouvelle génération (Gilissen, Hoischen et al. 2011), dû au contenu habituellement riche en GC et à la petite longueur des fragments séquencés.

En conclusion, jusqu'à maintenant, les stratégies utilisées dans le projet LOCA n'ont pas mené à l'identification du gène causal. La validation génétique des variants candidats ou l'envoi d'individus supplémentaires au séquençage permettra probablement l'identification du ou des gènes responsables de LOCA. L'identification de gènes causant des maladies est le premier pas vers la compréhension de ces maladies.

Bibliographie

Anheim, M., C. Tranchant and M. Koenig (2012). "The autosomal recessive cerebellar ataxias." N Engl J Med **366**(7): 636-646.

Bernard, G., I. Thiffault, M. Tetreault, M. L. Putorti, I. Bouchard, M. Sylvain, S. Melancon, R. Laframboise, P. Langevin, J. P. Bouchard, M. Vanasse, A. Vanderver, G. Sebire and B. Brais (2010). "Tremor-ataxia with central hypomyelination (TACH) leukodystrophy maps to chromosome 10q22.3-10q23.31." Neurogenetics **11**(4): 457-464.

Bertini, E., V. des Portes, G. Zanni, F. Santorelli, C. Dionisi-Vici, S. Vicari, G. Fariello and J. Chelly (2000). "X-linked congenital ataxia: a clinical and genetic study." Am J Med Genet **92**(1): 53-56.

Boycott, K. M., S. Flavelle, A. Bureau, H. C. Glass, T. M. Fujiwara, E. Wirrell, K. Davey, A. E. Chudley, J. N. Scott, D. R. McLeod and J. S. Parboosingh (2005). "Homozygous deletion of the very low density lipoprotein receptor gene causes autosomal recessive cerebellar hypoplasia with cerebral gyral simplification." Am J Hum Genet **77**(3): 477-483.

Bras, J., R. Guerreiro and J. Hardy (2012). "Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease." Nat Rev Neurosci **13**(7): 453-464.

Campuzano, V., L. Montermini, M. D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli, F. Zara, J. Canizares, H. Koutnikova, S. I. Bidichandani, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P. I. Patel, S. Di Donato, J. L. Mandel, S. Coccozza, M. Koenig and M. Pandolfo (1996). "Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion." Science **271**(5254): 1423-1427.

Carr, I. M., C. P. Diggle, N. Touqan, R. Anwar, E. G. Sheridan, D. T. Bonthron, C. A. Johnson, M. Ali and A. F. Markham (2012). "Identification of autosomal recessive disease loci using out-bred nuclear families." Hum Mutat **33**(2): 338-342.

Carr, I. M., K. J. Flintoff, G. R. Taylor, A. F. Markham and D. T. Bonthron (2006). "Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families." Hum Mutat **27**(10): 1041-1046.

Carr, I. M., C. A. Johnson, A. F. Markham, C. Toomes, D. T. Bonthron and E. G. Sheridan (2011). "DominantMapper: rule-based analysis of SNP data for rapid mapping of dominant diseases in related nuclear families." Hum Mutat **32**(12): 1359-1366.

Carr, I. M., S. A. Markham and S. D. Pena (2011). "Estimating the degree of identity by descent in consanguineous couples." Hum Mutat **32**(12): 1350-1358.

Casals, F., A. Hodgkinson, J. Hussin, Y. Idaghdour, V. Bruat, T. de Maillard, J. C. Grenier, E. Gbeha, F. F. Hamdan, S. Girard, J. F. Spinella, M. Lariviere, V. Saillour, J. Healy, I. Fernandez, D. Sinnett, J. L. Michaud, G. A. Rouleau, E. Haddad, F. Le Deist and P. Awadalla (2013). "Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans." PLoS Genet **9**(9): e1003815.

Delatycki, M. B., R. Williamson and S. M. Forrest (2000). "Friedreich ataxia: an overview." J Med Genet **37**(1): 1-8.

Doi, H., K. Yoshida, T. Yasuda, M. Fukuda, Y. Fukuda, H. Morita, S. Ikeda, R. Kato, Y. Tsurusaki, N. Miyake, H. Saitsu, H. Sakai, S. Miyatake, M. Shiina, N. Nukina, S. Koyano, S. Tsuji, Y. Kuroiwa and N. Matsumoto (2011). "Exome sequencing reveals a homozygous SYT14 mutation in adult-onset, autosomal-recessive spinocerebellar ataxia with psychomotor retardation." Am J Hum Genet **89**(2): 320-327.

Duquette, A., K. Roddier, J. McNabb-Baltar, I. Gosselin, A. St-Denis, M. J. Dicaire, L. Loisel, D. Labuda, L. Marchand, J. Mathieu, J. P. Bouchard and B. Brais (2005). "Mutations in senataxin responsible for Quebec cluster of ataxia with neuropathy." Ann Neurol **57**(3): 408-414.

Engert, J. C., C. Dore, J. Mercier, B. Ge, C. Betard, J. D. Rioux, C. Owen, P. Berube, K. Devon, B. Birren, S. B. Melancon, K. Morgan, T. J. Hudson and A. Richter (1999). "Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS): high-resolution physical and transcript map of the candidate region in chromosome region 13q11." Genomics **62**(2): 156-164.

Fogel, B. L. (2012). "Childhood cerebellar ataxia." J Child Neurol **27**(9): 1138-1145.

Gilissen, C., A. Hoischen, H. G. Brunner and J. A. Veltman (2011). "Unlocking Mendelian disease using exome sequencing." Genome Biol **12**(9): 228.

Griffiths AJF, M. J., Suzuki DT, Lewontin RC, Gelbart WM (1993). An Introduction to Genetic Analysis 5th Edition. New York.

Gros-Louis, F., N. Dupre, P. Dion, M. A. Fox, S. Laurent, S. Verreault, J. R. Sanes, J. P. Bouchard and G. A. Rouleau (2007). "Mutations in SYNE1 lead to a newly discovered form of autosomal recessive cerebellar ataxia." Nat Genet **39**(1): 80-85.

Hersheson, J., A. Haworth and H. Houlden (2012). "The inherited ataxias: genetic heterogeneity, mutation databases, and future directions in research and clinical diagnostics." Hum Mutat **33**(9): 1324-1332.

Jayadev, S. and T. D. Bird (2013). "Hereditary ataxias: overview." Genet Med.

Kruglyak, L., M. J. Daly, M. P. Reeve-Daly and E. S. Lander (1996). "Parametric and nonparametric linkage analysis: a unified multipoint approach." Am J Hum Genet **58**(6): 1347-1363.

Laberge, A. M., J. Michaud, A. Richter, E. Lemyre, M. Lambert, B. Brais and G. A. Mitchell (2005). "Population history and its impact on medical genetics in Quebec." Clin Genet **68**(4): 287-301.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Liu, C. X., Y. Li, L. M. Obermoeller-McCormick, A. L. Schwartz and G. Bu (2001). "The putative tumor suppressor LRP1B, a novel member of the low density lipoprotein (LDL) receptor family, exhibits both overlapping and distinct properties with the LDL receptor-related protein." J Biol Chem **276**(31): 28889-28896.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

Orr, H. T., M. Y. Chung, S. Banfi, T. J. Kwiatkowski, Jr., A. Servadio, A. L. Beaudet, A. E. McCall, L. A. Duvick, L. P. Ranum and H. Y. Zoghbi (1993). "Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1." Nat Genet **4**(3): 221-226.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.

Scharfe, C., P. Zaccaria, K. Hoertnagel, M. Jaksch, T. Klopstock, R. Lill, H. Prokisch, K. D. Gerbitz, H. W. Mewes and T. Meitinger (1999). "MITOP: database for mitochondria-related proteins, genes and diseases." Nucleic Acids Res **27**(1): 153-155.

Sobel, E., J. C. Papp and K. Lange (2002). "Detection and integration of genotyping errors in statistical genetics." Am J Hum Genet **70**(2): 496-508.

Stitzel, N. O., A. Kiezun and S. Sunyaev (2011). "Computational and statistical approaches to analyzing variants identified by exome sequencing." Genome Biol **12**(9): 227.

Tetreault, M., M. Srour, J. Allyson, I. Thiffault, L. Loisel, Y. Robitaille, J. P. Bouchard and B. Brais (2011). "Founder mutation for alpha-sarcoglycan-LGMD2D in a Magdalen Islands Acadian cluster." Can J Neurol Sci **38**(5): 747-752.

Thiffault, I., M. J. Dicaire, M. Tetreault, K. N. Huang, J. Demers-Lamarche, G. Bernard, A. Duquette, R. Lariviere, K. Gehring, A. Montpetit, P. S. McPherson, A. Richter, L. Montermini, J. Mercier, G. A. Mitchell, N. Dupre, C. Prevost, J. P. Bouchard, J. Mathieu and B. Brais (2013). "Diversity of ARSACS mutations in French-Canadians." Can J Neurol Sci **40**(1): 61-66.

van Gaalen, J. and B. P. van de Warrenburg (2012). "A practical approach to late-onset cerebellar ataxia: putting the disorder with lack of order into order." Pract Neurol **12**(1): 14-24.

Xia, J., Q. Wang, P. Jia, B. Wang, W. Pao and Z. Zhao (2012). "NGS catalog: A database of next generation sequencing studies in humans." Hum Mutat **33**(6): E2341-2355.